THE NEW EXAMINER

BY

PHILIP BOSWOOD BALLARD

M.A., D.LIT.

AUTHOR OF "TEACHING THE ESSENTIALS OF ARITHMETIC," "MENTAL TESTS" "TEACHING THE MOTHER TONGUE," "THE CHANGING SCHOOL," "GROUP TESTS OF INTELLIGENCE," "FUNDAMENTAL ARITHMETIC," ETC.

NEW IMPRESSION

UNIVERSITY OF LONDON PRESS, LTD.

WAR-TIME ADDRESS:

ST. HUGH'S SCHOOL, BICKLEY, KENT

FIRST PRINTED . . . *October* 1923
Reprinted . . . . . *July* 1924
Reprinted . . . . . *June* 1925
Reprinted . . . . . *April* 1927
Reprinted . . . . *September* 1929
Reprinted . . . . *November* 1933
First Cheap Edition . . *September* 1934
Reprinted . . . . . *June* 1935
Reprinted . . . . *October* 1936
Reprinted . . . . . *April* 1941
Reprinted . . . . . *April* 1944

Printed in Great Britain for the UNIVERSITY OF LONDON PRESS, LTD.,
by HAZELL, WATSON AND VINEY, LTD., London and Aylesbury.

PREFACE

THE mental tester shares the fate of all reformers in being sometimes misunderstood; and he shares the weakness of all mankind in preferring to be understood. And his duty jumps with his preference. For it is clearly his duty to define his aim and to explain his outlook. The schoolmaster is prone to regard him with suspicion. While there is much, he thinks, that is amusing about the mental tester there is also something of the sinister, and not a little of the dry-as-dust philosopher. He is at once too frivolous and too ponderous. The catches and conundrums which are supposed to constitute his stock-in-trade are well enough as a fireside diversion, but they have nothing to do with the serious pursuits of the school. And as for the weight of statistics that he brings to bear on the trivial replies of the children, it is using a sledge-hammer to crack a nut.

This, of course, is the veriest travesty—a complete misconception of the method and the purpose of the mental tester. For his purpose is simply to examine—to do what the ordinary examiner tries to do, but to do it better. He is out to explore the mind of each particular child, and to explore it with a thoroughness and a precision that have never been compassed before. He is in fact the old examiner in a new guise. His puzzles, when he has

any, are examination questions: they are tests, not jests. His statistical methods are the statistical methods of the old examiner extended and improved out of all recognition.

Rightly conceived, then, the mental tester is a reformer of examinations. That examinations need reforming is quite obvious. If the reader has any doubts on the matter they will at once be resolved by reading Mr. Hartog's book on *Examinations and their Relation to Culture and Efficiency*. In the pages of that excellent little book the fundamental weakness of current methods of examining lies open for all to see. Their weakness consists in their uncertainty—a capriciousness in the working and a lack of steadiness and accuracy in the results. To make examinations less precarious—less dependent on human variability—is the aim and purpose of all schemes of reform. With this end in mind methods old and new are being tried in the fire of experiment and statistical criticism; and it is curious to observe that the methods which come out best derive from quite humble sources. For, in its general structure and its technique of marking, the new examination more closely resembles the informal examination held by the teacher in his own class-room than it resembles the formal examination held in an examination hall under the auspices of an examining body. Tests that seemed to be the most careless and casual have turned out to be the most scientific and precise.

I have said that many see in the new tests a trace of the sinister—a hint of menace—a something that might be used to the disadvantage of the teacher and the detriment of his craft. Standard

tests bring to mind the old "standards of examination" of the government code; they suggest a screwing up of all scholars to a common level of efficiency, in spite of the tremendous differences stamped upon them by nature and by nurture. In suggesting this they suggest a perverted purpose. The norms of standard tests are meant for standard minds. To use the tests in isolation—to use them apart from other sources of knowledge—is to use them illegitimately. (They are designed for the schoolmaster, who knows his pupils; not for the critic, who knows them not.) They are intended as a special guide, not as a universal gauge; as a measure of what may be expected in some circumstances, not as a measure of what may be expected in all circumstances. And against the unfair use of standard tests of attainments the teacher has in intelligence tests a protection and a defence; for they enable him to exhibit the quality of the material on which he works.

Having said so much I must now admit the justice of the main charge: I must admit that the new examination is dangerous. It is dangerous because it is vital and forceful and efficient. Ideas themselves are dangerous—as dangerous as dynamite. Every agency, every power, every invention that can be used as a boon can also be used as a bane. The same discovery that made the ploughshare possible also made possible the sword and the javelin. This bi-polarity appears in all things. When the House of Commons discusses the finance of education and armaments, we are reminded that the money that should go to putting brains in is generally spent in blowing brains out. But money

is a good thing for all that. And so is the new examination. Its possible abuse by the injudicious is a risk which is worth facing, in view of the certainty of a beneficent use in the hands of the wise.

It is well that I should point out how this book differs from its two predecessors, *Mental Tests* and *Group Tests of Intelligence*. In the former books I was mainly concerned with standard tests of intelligence and of scholastic proficiency: in this book I am mainly concerned with the new methods of testing taken as a whole, and with the pressing of these new methods into the service of old purposes. I try to show how the technique which has proved of value in the testing of pupils' intelligence will also prove of value in the testing of pupils' progress. And the testing of pupils' progress is an indispensable part of the teacher's task. Indeed, the only kind of examination that is quite free from suspicion—the only kind that has never been called into question—is the kind that closely follows the course of study. It is the kind I had in view when I devised the examinations that appear in this book. They are meant as concrete examples of the principles of testing which I try to expound in the earlier part of the book; they are meant as models (imperfect models it is true, but still models) upon which the teacher may base his own class examinations; and, finally, they are meant as standard tests which may be used without change in schools of any kind.

Their use as standard tests needs a word of justification. If it is a sound theory that an examination which tests attainments should be moulded to fit

a scheme of work, what particular scheme of work do these examinations of mine fit? The answer is, no one particular scheme, but rather a general or skeleton scheme which stands for the highest common factor of a multitude of separate schemes. It represents the residue when all differences have been taken out. It is also, in intention if not in fact, the residue of knowledge that remains in the mind of the ordinary man when he has left his school days far behind, and the hand of time has thinned out his crop of textbook lore. While, for instance, there are certain geographical facts which every man with any pretence to culture knows (or ought to know), there are other facts of which he may justly, or even boastfully, be ignorant. No man need be ashamed of not being able to give the area of Madagascar or the height of Popocatepetl; nor need he be ashamed if he could not locate Macgillicuddy's Reeks; or, indeed, if he were a little uncertain in his mind whether the term referred to a geographical feature or to an article of clothing. He ought, however, to be thoroughly ashamed of himself if he did not know where Manchester was—if only to avoid going there. In constructing my typical examinations the plan I have adopted is to ask about Manchester and leave Macgillicuddy's Reeks alone. The facts and processes upon which the questions bear cannot possibly control a syllabus of work, for any syllabus that does not include them is beyond all control.

My main plea is that all reform of examinations moves in the same direction: it moves from the uncertain to the certain, from the subjective to the objective, from the precarious and personal to the

independent and impersonal—in a word, from guesswork to measurement. And in defending the thesis that all real examination is measurement I have plunged into metaphysics. It is a brief plunge, and a shallow; but a plunge that seemed to me to be necessary in order to meet the objections that are felt to measuring the workings of the human mind. (For it is the felt objections that are formidable, not the spoken.) The spectres to be laid are the doubts which emerge from such depths of the objector's mind that he feels their presence and their pressure without being able to define their shapes or to give them a local habitation and a name. And in dragging to the light of day my own inarticulate doubts (on the somewhat unsafe assumption that they are the same as those of my readers) in order to render them articulate and rational, or to force them to reveal their irrationality, I found it difficult to stop short of those first principles which belong to the realm of philosophy. And since in this realm to talk nonsense without knowing it is easier than anywhere else, I have taken the precaution to submit the chapters concerned to Dr. William Garnett and Professor Percy Nunn, to whom I am deeply indebted for their generous help and their illuminating criticism. They are in no way responsible, however, for what faults or weaknesses may appear in those chapters. In my blunders I claim to be wholly original.

My acknowledgments are due to the Editor of *The Times Educational Supplement*, who, with his customary kindness, has allowed me to use for some of the introductory chapters material which has

already appeared in his journal. I am also deeply indebted to the many teachers who have interested themselves in my tests and have supplied me with valuable data. Particularly am I grateful to Mr. J. G. Robson and to Mr. T. H. Elliott for the help which they have so readily and so liberally given. Finally, I wish to thank my colleague, Mr. John Brown, for his friendly service in correcting the proofs.

P. B. BALLARD.

CHISWICK, *July* 1923.

CONTENTS

THE NEW EXAMINER

CHAPTER I

ORIGINS AND PRINCIPLES

EXAMINATIONS began a long time ago. According to Mr. Hartog the first examiner was a Chinaman, and the first examination a civil service examination. This may be true, or it may be merely an instance of the general tendency to ascribe to the Chinese the invention of everything whose origin we cannot trace. Be that as it may, examinations for the few have existed throughout historic times,[1] whether they took the form of trials of skill, or of professional tests, or of the public disputations of which we read so much in the history of the Middle Ages, or of the formal written examinations of the present day. Examinations for the few have always been with us; but examinations for the many are a modern institution. In our own little island it was not until after the Crimean War—not until the appalling incompetence of the War Office officials became a

[1] Dr. Rusk, of St. Andrews, sends me the following curious passage which he found among the papers of the late Professor Edgar: "In ancient communities there was an official examiner, who, by severe and searching tests, examined or tried men's characters. Satan in the book of Job and in the Gospels is the inspector and tester of the apparently good man."

public scandal—that examination took the place of nomination as the ordinary means of entry to the Civil Service.

It was not, indeed, till the nineteenth century was well advanced that examinations were taken quite seriously by the English nation. In the eighteenth century they seem to have been of little account, even at the universities. John Scott, Earl of Eldon, who went to Oxford in 1766, testifies thus: "An examination for a degree at Oxford was a farce in my time. I was examined in Hebrew and in history. 'What is the Hebrew for the place of a skull?' I replied, 'Golgotha.' 'Who founded University College?' I stated (though, by the way, the point is sometimes doubted) that King Alfred founded it. 'Very well,' said the examiner, 'you are competent for your degree.'" This manifestly marks a point to which examinations had sunk, not a point to which they had risen. Nevertheless, there is no lack of evidence that in the eighteenth century examinations were in a bad way.[1] In the year 1784 there was published by "C. Dilly, in the Poultry," a book entitled, *Liberal Education, or a Practical Treatise on the Methods of Acquiring Useful and Polite Learning*, by the Reverend Vicesimus Knox, A.M. Knox was in advance of his times, for not only had he chapters "On Grammars and Introductory Books to the Latin," "On Making a Proficiency in Greek," and "On the Ornamental Accomplishments," but he also had chapters "On the Study of the English Language," and "On Learning to Speak," and finally "On the

[1] For further evidence see *The Older Universities of England*, by Albert Mansbridge, p. 104.

Utility of Examinations." He starts that chapter thus :

"Public examinations have of late been established in some colleges, and nothing has been found to contribute more to the success of the academic discipline. The same salutary consequences will flow from the practice, if it should be generally imitated, in the nurseries of the university."

Knox was an ardent advocate, not only of public examinations, but also of those less formal, but no less important, examinations which the teacher conducts in his own class-room.

If in the eighteenth century there was a "slump" in examinations, in the nineteenth century there was a "boom." The thing was overdone. Elementary education was hampered and shackled for many years by annual examinations conducted by the Government. Secondary education had its outlook narrowed and its aims warped by the necessity to prepare for certain competitive examinations which were supposed to set upon the schools the hall-mark of success. So keenly was the pressure of examinations felt that in November 1888 there appeared in *The Nineteenth Century* an article protesting in the strongest terms against our "sacrifice of education to examination," an article signed by 400 of the most eminent men and women of the day. From the thraldom of examinations externally imposed, elementary education is now almost completely emancipated. Not so secondary education. But even here the pressure is less than it was. Scholarships do not dominate the teaching so much as they did in the days gone by.

Here we catch a glimpse of some of the ways in

which the examination system may pass beyond its legitimate bounds and trespass upon the rights of the school. There are clearly two things which examinations should not do: they should not dictate the curriculum, and they should not be used as the sole gauge of the success of the teaching. They should not dictate the curriculum, for each school has its own soul to be saved in its own way; it has its own aims and its own ideals, to say nothing of the aims and ideals of its several pupils and its several teachers. It should be as much its privilege to develop its own curriculum as it is the privilege of a man to grow his own skin.

Examinations should not be used as the sole gauge of the success of the school or the success of the teaching; and this not merely because many of the finer products of the school are at present unexaminable, but because we cannot fairly judge the teacher's success without taking into account the native quality of the material upon which he has exercised his craft. In other words, an intelligence examination is necessary as well as an attainments examination. But the old system knows nothing of an intelligence examination. That is one of the inventions of the new examiner.

The new examiner came into being in this way. Towards the close of last century a few ardent educationists, seeing the need for a scientific study of the growing mind, started what was known as the child study movement. They made wide and searching inquiries, they bombarded home and school with *questionnaires*, they collected and collated a multitude of facts only a small portion of which will ever be published. It is rumoured

that in the store-rooms of Clark University, the university where Dr. Stanley Hall, the father of child study, is the presiding genius, there are tons and tons of documents awaiting the directive mind that will reduce them to order, or, as is more likely, the incendiary hand that will reduce them to ashes. For to the child study enthusiasts no trouble was too great, no item too trivial. They inquired into children's interests, their fears, their lies, their toys, the contents of their pockets, and the contents of their minds. No fault whatever could be found with the extent of the investigations; the quantity of the *data* was beyond reproach; the quality, however, was a little dubious. When, moreover, the researchers had got hold of the essential facts they did not know what to do with them. Nothing happened. No problem was solved, no light emerged to guide the teacher in his daily task. Let us take as an instance Dr. Stanley Hall's celebrated investigation into children's dolls. Impressive as was the avalanche of figures and diagrams that drove home the fact that a little girl's interest in dolls culminated at eight and a half years of age, it was not easy to see what that particular fact had to do with schooling. Certain it is that it had no effect on the curriculum. Nowhere in the wide world is there a school where it is part of the programme for girls of eight and a half to play with dolls.

It thus happened that early in the twentieth century child study began to pass into a new phase. It became experimental pedagogy. In its new guise it was pursued with vigour by such men as Thorndike in America, Binet in France, Meumann in Germany, Claparède in Switzerland, and Winch

in England. It had a number of clear and definite aims. It sought to solve certain fundamental problems that perplexed the administrator in his office and the teacher in his class-room. It endeavoured to find out what should be taught, when it should be taught, and how it should be taught. It was vastly concerned about the amount of transfer that took place from one line of mental activity to another line of mental activity. It tried to discover which branch of study fatigued the student the most, and on which day of the week and at which hour of the day he could put forth the most strenuous mental effort. It tried to find out which was the best way of getting a poem by heart: whether it was by studying it as a whole or by attacking it line by line. In fact, there was no end to the questions raised and the experiments projected. But on the whole the results were disappointing. Not that they were worthless—indeed some of them were of great and permanent value—but they did not reach that degree of validity which is supposed to be attained by the rigid application of scientific method. They fell short of complete certitude because the methods of evaluating the mental products lacked the necessary precision. They failed, in fact, when they did fail, because the investigators did not know how to examine. They knew how to examine better than the old examiner, because they were able to bring into the class-room mathematical methods learnt in the psychological laboratory. But even these were not sufficiently steady or sufficiently precise. Experimental pedagogy broke down, not in its pedagogy, but in its mathematics.

The consequence was a gradual shifting of interest from experiment to measurement. One by one the researchers abandoned their researches and turned their attention towards improving the means of research—towards the devising of scales for measuring mental processes and mental products. The defection is now almost complete; but it is, I need scarcely say, merely a temporary defection. Psychologists will return much better equipped for their task. They will return bearing in their hands a new system of pedagogical mathematics.

This transfer of professional interest forces itself on our notice when we contrast the earlier with the later numbers of journals devoted to the science of education. One of the best of these is the *Journal of Educational Psychology*, the first number of which appeared in January 1910. Among a large number of articles and reviews there was only one reference to problems of measurement. The last issue of the same journal scarcely touches any other problem. It is as difficult to find in it a page that does not deal with testing as it is difficult to find in the first number a page that does.

For the more strictly mathematical side of testing we can find no safer guide than Brown and Thomson's *Essentials of Mental Measurement*. The bibliography which appears as an appendix of that book has 207 references, and of these only 19 date earlier than the beginning of this century. No clearer proof is needed of the recent emergence of a new mode of procedure in the probing and the measuring of people's minds.

Three important milestones stand out on the road from the old to the new: (1) the publication

in 1897 of Dr. J. M. Rice's article in *The Forum* applying scientific measurement to school attainments; (2) the issuing in 1908 of Binet's Intelligence Scale; and (3) the setting of Intelligence Examinations in 1917 to the American Army. The first marks the beginning of standard tests of attainments, the second the beginning of standard tests of intelligence, and the third the beginning of group tests of intelligence.

So, whether we like it or not, the new examiner has arrived, and it would be well to look carefully at the sort of examination he sets.

The most fundamental feature of the new examination is that it is consciously and designedly a measurement. It is a measurement and not a guess. Nor yet an award. The old examiner is said to "give" or "award" marks, as though marks were a sort of prize or bounty the value of which depended as much upon the generosity of the giver as upon the merit of the receiver. The new examiner, with his fine spirit of scientific aloofness, would scorn to regard his scores and norms as anything but discoveries; they were there already, he merely brought them to light.

To reach the precision of measurement at which the new examiner aims a scale is necessary; and a scale that is finely graded. Rough scales with large units are almost useless. It is idle to measure a field with a scaffold-pole, or the works of a watch with a foot-rule. And the fine grading of the scale is reflected in the initial units—the questions set. While the old examiner sets a small number of large questions, the new examiner sets a large number of small questions. At a university exami-

nation about a dozen questions have to be answered in three hours; at the American Army Alpha examination 212 questions had to be answered in less than half an hour. It is clear that the new examiner atomises his test; discarding the single bullet, he charges his gun with buckshot.

The mode of marking at which the new examiner aims is purely objective: he tries to get a scale which will give the same score for the same paper whoever the examiner may be. (The Americans call it making the examination fool-proof.) A fool-proof test is one which prevents the examiner from making a fool of himself. It is generally believed that making a fool of oneself is a practice peculiar to the examinee; the examiner is supposed to be free. But this is because he was never found out. Until quite recently nobody thought of examining the examiner; but he himself has now been weighed in the balance and found wanting. His claims to reliability, to say nothing of infallibility, can no longer be maintained. And since marks that fluctuate with the caprice or the capacity of the examiner have no scientific validity, expedients have been found which take away his discretionary power and bind him down to a rigid scheme of scoring. To secure certitude in the marking, and thus not only disarm all suspicion of personal bias but also remove the sting of injustice from an adverse verdict, is well worth attempting, even at some sacrifice.

An illustration will make this clear. One of the questions at the American Army examination was: Why is beef better food than cabbage? If this were set as at an ordinary examination it would produce a varied crop of answers. A vegetarian

would begin by denying the fact, and would proceed to prove that beef was a slow poison, and cabbage a food which filled a man with vitality and vim. The gourmand, on the other hand, would argue that, properly speaking, cabbage was not food but fodder. And all would, if they could, pour forth platitudes about proteids and carbohydrates and vitamines until the poor harassed examiner would find himself distributing marks wildly and lawlessly. This waste of words and temper is prevented in the army test by the simple device of attaching to the tail of the question these possible answers, thus :

Because it tastes better.
Because it is more nourishing.
Because it is harder to obtain.

The candidate is merely required to put a cross in front of the best answer. And in order to make sure that no mistake is made in the scoring the examiner is told that the cross should appear before the second answer. That ties his hands and takes away all his capacity for mischief. The sacrifice is obvious. The scheme does not give our Chestertons a chance; nor does it distinguish between the literate and the illiterate, between the man of science and the ignoramus. But then it is not intended to do so; it is not a test of English, nor of science; it is merely a test of common sense. And it is only one out of 212—a little thing that does a little piece of work, but does it with regularity and with certitude.

Again, the new examiner standardises his tests before he puts them to serious use. At the public examination of to-day the paper set is always abso-

lutely fresh. Its contents have been kept a dead secret right up to the moment when the ordeal starts. The questions never having been used before, the examiner has merely a preconceived and unverified notion of their difficulty and of the kind of response they will evoke. He has no means of judging how the candidates compare with the total population; no means, or very imperfect means, of comparing them with candidates of previous years. The examination is isolated and self-contained; it fails to relate itself to other examinations or its candidates to other candidates. The new examination, on the other hand, stands in a known relation to other similar examinations; it is part of a complex network of standard tests. For the new examiner carefully tests his tests beforehand; he applies them to children in other parts of the country and submits the results to statistical analysis; he eliminates questions that are demonstrably weak and substitutes others that are more cogent; he grades his questions in order of difficulty, or in any other order that suits his purpose; he secures age norms and class norms and finds an index of the degree of dispersion; he compares his results with those obtained by other standard tests; and sometimes he gets out a parallel series of questions similar in type and equal in difficulty—a series which may, in case of need, be substituted for the original series. And all this can he venture to do because the evils of special coaching for the new examination are far less serious than the evils of special coaching for the old.

The old examiner bases his verdict on the evidence afforded by a single examination; the new examiner

is rarely satisfied with less than two. He tries first of all to discover the native intellectual gifts of the candidate, and then to discover what use has been made of those gifts. One branch of the dual examination appraises the pupil's mother-wit, the other branch appraises his book-learning. And if the examiner were forced to limit himself to one of these two types of tests he would almost invariably choose the former; for he has good ground for believing that, for most of the purposes for which public examinations are held, an intelligence test is more significant than an attainments test. It is more significant because it affords a clearer and more conclusive index of what the testee is likely to achieve in years to come. It is a promise of the future, and not merely an epitome of the past. True, the orthodox examination leaves neither of these factors out of account; but it fails to distinguish between them; it does not tell us how much of the result is due to nature and how much to books, tuition, and opportunity.

Another characteristic of the new examination is the small amount of writing demanded of the candidate. He is required to think, and sometimes to think furiously, but he is not required to write more than an occasional word or symbol. In the group tests of intelligence, so widely used in America, all the pupil has to do is to make a mark of some kind on his question paper—to underline a word, to encircle a number, or to indicate by a cross a selected phrase. And therein it stands in violent contrast with the orthodox examination, which demands from the candidate a great flow of words from the pen. Nearly every examination is an examination

in English : the answers consist of a series of exercises in English composition. And important as English composition is as an intellectual discipline, as a basis of measurement it is egregiously bad. It fails through its very wealth and complexity. It confuses the examiner by telling him too much, and diverting his attention to irrelevant merits and irrelevant defects. Be that as it may, the new examiner fights shy of such complex mental products as an essay or a drawing, because he has not yet been able to render them amenable to the finer methods of measuring. Professor Thorndike and others have tried to do it, but they have tried in vain.

It is now manifest to the reader that the general trend of the new system is to transfer the burden of care and thought from the shoulders of the marker of the test to the shoulders of the maker of the test. At present the maker has much the easier time. The man who sets an examination paper will " knock it off " in an hour or two, but the man who reads the scripts will have to toil over them for days, or, in the case of a big public examination, for weeks and months. And the money spent in marking the papers is incomparably more than the money spent in setting the test. The new tendency is to reverse all this ; and so much time and thought are devoted to the devising and revising and standardising of the test that the appraisal of the results may safely be left to subordinates. The bulk of the work is done before the candidate sets pen to paper. In fact the motto of the new examiner seems to be : Look after the testing and the marking will look after itself.

CHAPTER II

EXAMINATION AS MEASUREMENT

(*a*) The Thesis

Rarely have examinations been looked upon with a friendly eye. The common practice is to denounce them freely, and then to tolerate them as a necessary evil. They have been variously described as "an enemy of the pupil," "an obstacle to education," "a begetter of rivalry and strife," "an emetic," and "a glorification of memory at the expense of the rational and creative powers." Thompson, the late Master of Trinity, called them "a presumptuous attempt to gauge the depths of human ignorance." There has been, in fact, a strong tendency to regard examinations as essentially mischievous. The plain truth, of course, is that an examination—an ideal examination—is as neutral as the equator, as free from malice as a pint pot,[1] a foot rule, or a clinical thermometer. Its essence is simply measurement. The evils and dangers attendant upon it are the evils and dangers of bad measurement, and the only way to improve examinations is to improve our methods of measurement.

It is important to realise that measurement is concerned with magnitudes, that magnitudes are of different kinds, that the thing measured is always

Empty.

an abstraction, and that the man who measures should always use an objective scale and arrive at an objective finding.

The first characteristic of measurement is that it deals with magnitudes—with things that can be compared on the basis of "more or less." This is the very least condition that must be satisfied in order that mathematical laws may be applicable at all; the very least condition that will justify our saying:

I. That of two things of the same kind one must be either greater than, equal to, or less than the other.

II. That things that are equal to the same thing are equal to one another. And

III. That if the first of three things is greater than the second, and the second greater than the third, then is the first greater than the third.

These are axioms which pertain to all magnitudes; but there are other axioms which do not. The axiom that the whole is greater than its part, for instance, although it seems essential and fundamental, does not really apply to magnitudes of all kinds. It applies to extensive magnitudes but not to intensive magnitudes. It applies to the length of a piece of timber but not to the loudness of a musical note. Regarded from a psychical and not from a physical point of view, a note of a given loudness cannot be said to be a whole made up of other notes of different loudness. Nor will two notes of the same loudness sounded at once give a note of double the loudness. We cannot, in fact,

add loudnesses as we can add thicknesses; nor can we subtract them—not at least in the same way. The difference between one extensive magnitude and another is itself an extensive magnitude; but the difference between two intensive magnitudes is not an intensive magnitude but something else—something which we represent by a distance along a graduated line. Five yards of ribbon is greater than three yards of ribbon by two yards of ribbon; but the temperature of boiling water is greater than the temperature of freezing water not by some other temperature of water but by a number of degrees on the thermometric scale. Not only do extensive magnitudes differ in kind from intensive magnitudes but they differ in kind among themselves. The pitch of a musical note is a kind of intensity; and so is its loudness; but the physical basis of the one is very different from the physical basis of the other: one depends on the rapidity of the vibrations of the sounding body, the other depends on the amplitude of those vibrations. The common bond that brings them under the same rubric is that both admit of degrees, and the common defect that excludes them from the realm of extensive magnitude is that neither in the strictest sense consists of a whole which can be divided into parts.

The same thing is expressed by saying that everything in the world is either a quantity or a quality. A quantity (or extensive magnitude) can always be measured by a fixed unit of its own kind; but a quality (or intensive magnitude) cannot. It can often, however, be gauged by means of a scale which is purely conventional and artificial. It is im-

portant to note that in the progress of science items are being transferred from the category of quality to the category of quantity. There was a time when it seemed ridiculous to talk about heat as a quantity; it is by no means ridiculous now.

It will thus be seen that in things mental as well as in things physical there are magnitudes of different kinds, and these magnitudes are amenable in varying degrees to the principles and processes of arithmetic.

The next characteristic of measurement is that it deals with abstractions. Although we generally and legitimately speak of measuring a *thing*, we do not really measure a *thing*: we measure a single property of a thing—its length, its breadth, its area, its volume, its weight, its temperature, its cost, its hardness, or its durability. Indeed there is scarcely a limit to the number of attributes to which we may—some day if not now—assign a quantitative mark. But we must treat each attribute separately. We cannot measure two at a time, nor can we, without a further weighing of values, lump together the measurements of different attributes and find a single index for the lot. There is no symbol known to mathematics that can record at once the length and the weight and the cost of the pen with which I write. In dealing with material things we never dream of mixing categories in this way; but in dealing with things of the mind we are constantly doing it. We do it mildly in the marking of sums, flagrantly in the marking of essays; and to realise that we do this irrational thing is the first step towards reform.

Another essential trait of measurement is its objectivity. I do not propose here to discuss the variant views of the distinction between subjective

and objective. Carlyle describes Coleridge as talking through his nose about " sum-m-ject " and " omm-m-ject " ; and other men of letters have poured scorn upon these words. They are innocent words, however, and very useful. Common sense readily admits that in all knowledge there is something that belongs to the subject who knows, and something that belongs to the object that is known—something that is internal and something that is external —something that is private and personal and something that is common and universal—in fine, something that is subjective and something that is objective. If I am annoyed at receiving a letter which I consider abusive, my annoyance is subjective, the letter is objective, and its abusiveness may be either the one or the other. The personal element, especially when it distorts one's estimate of the objective reality, finds popular recognition in the hackneyed term " personal equation," a term borrowed from a science where the measurements have to be meticulously exact—the science of Astronomy. So exact and accurate have they to be that certain technical devices are adopted to detect the personal equation and to cast it out of the reckoning.

Quite apart, however, from measurement proper—measurement which deals with one single quality, attribute, or property, and is solely concerned with different amounts or degrees of the same quality—there is a distinct type of measurement which deals with two or more qualities, attributes or properties, and is concerned with a relationship between qualities that are different in kind. If, for instance, I say that one parcel is heavier than another, I am

comparing the two in respect of one particular attribute (weight), and asserting that one parcel has more of that attribute than the other; or if again I assert that John can spell better than James, I make a parallel statement respecting an attribute which is called "ability to spell." When, however, I say that in appraising diamonds, purity is to be regarded as more important than size, or that in marking arithmetic, accuracy is to be regarded as more important than neatness, I am making quite a different sort of judgment. Instead of measuring weight with a unit of weight and spelling ability with a unit of spelling ability, I take the results of measuring weight and purity and accuracy and neatness, and measure these particular findings with units which are as different from the original attributes as anything possibly can be. I measure them with units of *value*. In fact, I am measuring a magnitude of a peculiar kind—one which is highly abstract and which seems to belong less to the thing measured than to the man who measures. As a mark or measure value seems less stable and permanent than the primary elements upon which it is based; for value obviously depends on the purpose of the valuer. To an accountant the relative values of arithmetic and spelling are very different from what they are to a typist. One puts arithmetic before spelling and the other spelling before arithmetic.

But whatever type of measuring we use, whether we measure or evaluate, whether we use a pint pot, a foot rule, or a list of prices, we are fundamentally doing the same thing. We are referring new things to an old scale. The scale is always there. It

may be as objective as that of the draper who serves a customer with a yard of calico, or as subjective as that of the magistrate who gives one man seven days and another man three months for the same offence. Each has either his scale of measurement or his scale of values to which he can point if his findings are challenged.

It is important to note that in all complete and accurate measurement—when we take the thing as a whole as distinct from its several attributes—when, in fact, we fully appraise or assess anything, both types of measurement are necessarily involved. I cannot say all that is to be quantitatively said about even a piece of timber by simply measuring it in one way only, or indeed in a hundred ways; I must compare and evaluate the results of each of the separate measurements. Truly to assess its worth for any practical purpose of life, whether for firing, for building, or for furniture, I must evaluate it as well as measure it; and evaluate it separately for each purpose. And here, indeed, comes the great difficulty in appraising a product of the human mind. If it were only a matter of measuring it along one particular line by the application of a scale universally recognised and universally valid, the task would be comparatively easy. But it means much more than this. A picture, for instance, may be studied and rated from different points of view; we may consider its faithfulness to nature, the strength of the drawing, the beauty of the lines, the arrangement of light and shade, the balance of its parts, the harmony of the colours, or, finally, that impalpable and elusive quality upon which its æsthetic appeal essentially depends. And

we have left out of account such external and adventitious things as the reputation of the artist and the public demand for that particular kind of picture. To measure all these qualities separately is no easy task: to evaluate them afterwards seems well-nigh impossible. For it involves a scale of values, an arrangement of all the various qualities in a sort of hierarchy which not only says that one is higher than another but by how many units of value it is higher. Does such a scale exist—exist, that is, as an objective fact? Is it not true to say that each man has his own private scale of values and that these private scales are quite independent of one another? If the answer to this question is "Yes!" and no objective scale of values exists or can exist, then examinations are indeed in a bad way; and, more serious still, all attempts to improve them are hopeless. For the degree to which examinations are independent of the personal and private opinions of the examiners is the degree to which the results are scientifically valid. But there are strong reasons for believing that there *are* objective scales of value, reasons which will be given in due course.

CHAPTER III

EXAMINATION AS MEASUREMENT

(*b*) OBJECTIONS

THERE are some teachers—not many—who take up an antagonistic attitude towards modern methods of testing. They boldly assert that measurement has no place in education; for education deals with things of the mind—with impalpable, imponderable things, such as tendencies and interests and aspirations and ideals—things that have no more to do with measuring-tape and foot-rule than has the fragrance of a rose or the beauty of a star. The obvious reply is that the very people who make this assertion do, as a matter of fact, daily and hourly measure the things they so confidently proclaim to be immeasurable. But they do it carelessly and casually and vaguely. They say that Richard writes a better hand than Tom, or that Mary is more intelligent than Ruth. They declare James to be beyond his years in mathematical skill. They call one essay good, another fair, and another excellent. They range their pupils in order of merit, they praise them and they blame them, and they emphasise their estimates by rewarding the good and punishing the bad. Above all, they examine. No school is run or ever has been run without examinations. There may be no external

examinations, there may be no competitive examinations, but examinations of some sort, formal or informal, there must be if education is to go on at all. For the aim of education is to bring about certain changes in the pupil, changes which are in themselves invisible; and unless some means are taken to render them visible—and examinations are the best means yet devised for this purpose—the teacher gropes about in the dark. The objectors to measurement in education must, therefore, if they are consistent and serious, either give up examinations altogether and cease to express even opinions on the merits and attainments of their pupils, or else take up the strange position that examinations are like puns: the worse they are, the better they are—the less they are concerned with evidence studiously gathered and carefully verified the more fully do they serve the higher purposes of education. The bulk of teachers, however, clearly see the illogicality of that position: they realise that tests and examinations are an essential part of the educational system, and that the right thing to do is not to end them but to mend them. And to mend them is what the modern experimental psychologist is strenuously trying to do.

But many earnest people look coldly on the psychologist's efforts because they believe that education has to do with spiritual forces which are in their nature infinite, which belong to a realm where the categories of time and space have no meaning and where distinctions between large and small and between high and low do not exist. St. Paul, however, did not think like this: from

the spiritual world he did not exclude magnitudes and values. "And now abideth faith, hope, charity, these three; but the *greatest* of these is charity." There are degrees everywhere, even as "one star differeth from another star in glory." If we could transcend the differences between great and small, or high and low, if by some metaphysical alchemy we could achieve that higher synthesis of which the Hegelians so persistently speak, we should at the same time transcend all human experience and get beyond that work-a-day (and play-a-day) world, with which alone this book is concerned. To confuse here these categories and to revel in the vague and illimitable seem more likely to make a man muddle-headed than to make him spiritually-minded. There is nothing specially noble in the vague; there is nothing specially moral in the indefinite. The witness who said that the stone thrown by the prisoner was as big as a lump of chalk was no better as a man than as a witness. A man cannot claim to be good on the ground that his mathematics are bad. The vision seen by St. John the Divine had in it nothing of the nebulous. The walls of the New Jerusalem were not of illimitable height: they were precisely a hundred and forty-four cubits.

Nor do metaphysical doctrines of the ultimate value of time and space help us out of our difficulties. To quote Oliver Wendell Holmes: "Curious entities, or non-entities, space and time! When you see a metaphysician trying to wash his hands of them, and trying to get rid of them so as to lay his dry, clean palm on the absolute, does it not remind you of the hopeless task of changing the colour of

the blackamoor by a similar proceeding? For space is the fluid in which he is washing, and time is the soap which he is using up in the process, and he cannot get free from them, until he can wash himself in a mental vacuum." But in spite of the speculations of non-Euclidian geometers and of the modern apostles of Relativity, the ordinary man finds it far easier to think of space as infinite than to think of it as finite. And yet, if anything at all is measurable, space is measurable. And so is time. It is readily admitted that Love, as a Christian virtue, is infinite; but we can nevertheless love one person more than another. And if we take Love as a type of things spiritual, when we say that a spiritual thing is infinite all we practically mean is that we cannot monopolise it, nor economise it, nor exhaust it: in getting more of it, we do not necessitate somebody else having less; in taking less of it we do not thereby leave more of it for other people. It is not a commodity which comes under the laws of supply and demand. The amount of it that any man may receive is limited solely by his capacity to receive it. But it is limited by that; and being limited by that, it becomes amenable to measurement.

Most of the objections that have been urged against pressing the notion of measurement into the service of education are found on close investigation to be groundless. There is one, however, that is not. It is the difficulty of finding an objective scale. Without an objective scale measurement would be meaningless. For if each man had a private scale of his own, his quantitative judgments would have no significance for anybody but himself,

and no validity even for himself. They would belong to the realm of whims and fancies rather than to the realm of demonstrable truths. They would indeed be "merely a matter of taste"—the kind of taste about which it has truly been said there is no disputing. If taste, in fact, is purely subjective—if it has no objective element at all—not only can we not argue about it: we cannot even talk about it. Having no factors common to the experience of mankind it affords no true coinage for the currency of thought. We cannot discuss colours with the blind, nor music with the deaf. And with the man who believes himself to be Julius Cæsar, or a crocodile, or the village pump we refuse to argue, because we consider him to be out of touch with the universe of orderly human experience.

It is clear, however, that in that simple form of measurement which I have called measurement proper an objective scale is always possible. For what we directly measure is itself objective, and what we measure with is made up of the same kind of stuff. We measure length with length and weight with weight and spelling with spelling and sums with sums. In dealing with things of the mind it is mental products—overt, tangible products that we actually measure—and only inferentially do we measure the mental processes that are behind these products. It is possible to devise a test in simple arithmetical processes, and to provide a scale for marking it, which is as free from all suspicion of unfairness, or of personal bias, as the balance in the chemist's laboratory. Thus it happens that, in measurement proper, it is easy to convince ourselves that objectivity is secured. When, however, we

come to evaluation, difficulties meet us on every hand.

At this point let us digress a little to attain a clearer view of what is meant by objectivity. The distinction I am trying to draw between the subjective and the objective has little or nothing to do with one's belief about the ultimate reality of things.

The philosopher tries with his dissecting knife to cut a man's experience into two parts, and to refer one part to the man's mind and the other part to an independent universe with which that mind comes into contact. But different schools make the division in different places. The extreme realist leaves nothing to the mind but its own activity and its own manifest products—its feelings, volitions, fantasies, and dreams. Others (like Locke) place secondary qualities of things such as colours, sound, and taste, inside the mind, and the primary qualities of things, such as form and size, outside the mind. The extreme idealist refuses to make the cut at all; and holding that, in its ultimate analysis, all experience is mental, he regards the whole universe as spun out of some sort of consciousness. "All we see and seem is but a dream within a dream." The curious thing about these various theories is that none of them can be disproved. Each fits the facts, and no crucial experiment can be devised which will rule out all theories but one. But this does not matter here, for whichever theory we hold, there will still be a difference between a fancy and a fact, a phantom and a reality, a wrong solution to a problem and a right solution to a problem.

We must rid our minds of the idea that the

objective world is merely the world of material objects. Material objects form part of that world, but only a part. The rest is made up of relations between things, laws, principles, truths, "universals"—ideas if you will; but ideas in the Platonic sense, ideas which stand outside the individual mind, and which the individual mind "knows," but does not create. It is a world that cannot be changed at the mere caprice of the thinker. On the contrary, it exercises upon him that peculiar constraint which Stout makes the ground of distinction between belief and imagination. And the constraint is the same whether it is exercised by a material object or an abstract truth. I can no more believe that two and two make five than I can believe that the desk at which I write is a camel. The first is just as hard and inexorable a fact as the second; and just as truly a part of the objective world.

The numerical statement which I have just given as an instance of abstract truth is peculiarly convincing. Nobody doubts that mathematical facts belong to the objective world; nobody doubts that the natural number series (one, two, three, etc.) is found by the individual mind and not made by the individual mind; nobody doubts that the spatial relations revealed by the study of geometry are real relations belonging to a world of knowledge which is open to everybody and is the same for everybody; in fine, nobody doubts the objectivity of mathematics, of mathematical scales, and of mathematical measurements.

When, however, we enter the domain of morals doubts begin to assail us. We wonder whether there

is a scale of morals which is objective in the sense that the multiplication table is objective. We may be quite ready to believe that "it is a sin to steal a pin, much more to steal a greater thing," but we are not so ready to believe that the wickedness of stealing is directly and exactly proportional to the market-price of the thing stolen. Still less ready are we to believe that we can compare and evaluate in mathematical terms different kinds of wickedness, and say, for example, how many thefts are equal to one murder. And our scepticism does not arise solely from the fact that morality is more concerned with the inner motive than with the overt act: it springs also from the uncertainty that inheres in the scale of moral values. It sometimes goes so far as to deny that there is an objective scale at all. Of the many inconvenient facts that have to be explained away before the existence of such a scale can be demonstrated, none is more disturbing than the apparent dependence of moral standards on accidents of time and place. Morals seem to be almost a matter of geography—or of history. The scale of values of the Nietzschean philosophy is poles apart from the scale of values of the Christian philosophy. The "magnificent" man of Aristotle's ethics would nowadays be looked upon as a prig, a snob, and a pompous humbug. Facts of this kind give rise to the suspicion that morals, instead of standing staunch and immutable among the eternal verities, fluctuate like the fashions, and shift and change with race and clime. It has indeed often been pointed out that there is scarcely a vice known to humanity which has not somewhere or at some time been regarded as a virtue, and

scarcely a virtue which has not somewhere or at some time been regarded as a vice.

But uncertain as is our ground in the realm of morals it is more uncertain still in the realm of art. And many who readily accept moral values stoutly refuse to accept æsthetic values. For æsthetic values seem to be entirely subjective and relative. Personal taste is the only arbiter. What to one man is a beautiful picture is to another a ghastly horror. One school of painters tells us that the art of painting started with Cézanne; another school assures us that it ended with him. Even when the scale of values is obviously not a personal one it is at best a local or tribal one, and lacks the clear validity of universal truth. Darwin reminds us that Red Indians regard hair on the face as a deformity, and that the Chinese canon of taste in personal beauty demands that noses should be flat. As in painting so in music. What is one man's music is another man's noise. Everywhere in the domain of art do we find likes and dislikes, opinions and prejudices, violent assertions and violent denials; but nowhere do we find that calm acceptance of facts and principles which is so fine a characteristic of the world of mathematics.

CHAPTER IV

EXAMINATION AS MEASUREMENT

(*c*) The Reply

How are we to escape the quagmire of doubt and uncertainty into which we are forced by the above reflections? For escape is devoutly to be wished. If there are no general canons of taste and criticism in the matters with which education is most vitally concerned, if each man is here a law unto himself, with his own private scale of values which stands in no necessary relation to other scales, whose opinion is it that should prevail when concerted or communal action becomes imperative? If the multitude of private and personal scales of value are all of equal worth, then rational selection is impossible; if some are better than others, how are we to discover which are the better? We should need another and independent scale of values to do this. And where is this scale to be found? Is this, too, a private scale? The further we press the inquiry the clearer it becomes that a logical and consistent theory of the subjectivity of mental, moral, and æsthetic values is absurd and unworkable. Its advocates have to take refuge in an appeal to experience. They say that the man who has most thoroughly familiarised himself with the matters in question is the "expert"—the

experienced one—and it is his opinion that counts; it is his opinion that should prevail over all others. But does not this imply that his opinion is nearer the truth? And how can it be nearer the truth unless experience brings the mind into juster relationship with external values? Were it not indeed that experience brings about an adjustment between the self and something that is not the self, experience would be wholly vicious. For the only other thing it could do would be to intensify the idiosyncrasies of the self. As a matter of fact, it does both of these things. And this double function should be borne in mind by those who glorify experience at the expense of initiative and vision. They should remember that experience tends to stereotype as well as to illuminate; that in furnishing the mind it reduces the elbow-room. If it increases the number of ideas, it often blunts the capacity to deal with those ideas. If, again, it does not give a wider scope for thought but merely more familiarity with the same ideas—then by deepening the groove in which the mind runs it necessarily narrows the field of view. Whichever way we look at it, experience is seen to give with one hand and to take away with the other. The most we can do is to see to it that the giving is greater than the taking away. Unless there are external values with which the experiencing mind gets into closer and closer relationship—if there are no external and impersonal values—then a man is the worse rather than the better for being an expert. For in this case the less experienced a man is the wider and more generous is his outlook.

But I accept neither the doctrine of the essential

subjectivity of values, nor the doctrine of expert or experienced opinion. Both doctrines are, to my mind, utterly vicious. And, against them, I put forward what I regard as the true faith; which is this. The objective world, the world with which every individual mind, according to its capacity, comes into partial and changing relationship, consists of an indefinite multitude of entities, material and immaterial, concrete and abstract, some of which are spatial and temporal, others of which are both spaceless and timeless but are nevertheless essential parts of the objective whole. The one attribute common to all of them is that they exercise a controlling and constraining influence upon all human minds that come into vital relationship with them—an influence which tends to uniformity of human beliefs and community of human knowledge. This tendency (for it is a tendency only) may be deflected or disturbed (and generally is deflected or disturbed) by certain peculiarities on the part of the individual mind. It does not matter whether this uniformity of belief, this community of knowledge, is due to the fact that there is one universal thinker (or dreamer if you will) thinking (or dreaming) through all the several minds at once, or whether there is a real world of "things in themselves" outside the individual minds, which things themselves exercise the constraint towards uniformity. The constraint is there, wherever it comes from. And the extent to which the individual mind can rid itself of subjective resistance and yield to the constraint is the extent to which it arrives at truth. Among the many relationships that hold in the objective world, not the least manifest are magni-

tude, degree, and value; and if these relationships are there, it is well to believe that it is not beyond the wit of man to devise means of measuring them.

The progress of every science has been a progress towards objectivity and exactitude. The science is gradually purged of what is private and personal and brought nearer and nearer to the precision and certitude of mathematics. Beginning with casual observation, it ends with exact measurement; beginning with opinion, it ends with proof. And the controversy that rages round the question, Is education a science? can never reach a clear issue till it is put in another form, namely, How far has education passed along the road from the casual to the precise? In other words, To what extent is education an *exact* science?

Let us now return to the apparent instability of moral and æsthetic values. Everyone has his own sense of values, which is necessarily subjective; but the point at issue is whether there is not also an objective scale of values which is the *fons et origo* of all the individual scales, and to which these individual scales tend to approximate. The thesis I wish to establish is that there always is such an objective scale, and that the wide variation in individual scales in no way disproves its existence.

Before the invention of the thermometer (the thermometer is no older than King James's Bible) estimates of temperature were almost purely subjective. People judged by their feelings of heat and cold. A warm day to one would be a cool day to another. And if they quarrelled about it there was no external court to which they could appeal. They could only appeal to the feelings of

a third person no less fallible than themselves. Nor was it possible, except in the roughest way, to compare, in respect of temperature, one country with another country, or one period with another period. A sceptic might in those days, if he had a turn for dialectic, make out a plausible case for the pure subjectivity of temperature. He could not do so to-day. A simple little thermometer would shatter his arguments to pieces.

Another example, more significant still. If we present to a class of children about four years of age a simple line-drawing of a five-pronged fork or rake, and ask them to copy it, in nine cases out of ten they will exaggerate the number of prongs. The average number will be at least seven. Apart, that is, from the wide variations in individual judgment there will be a constant error of two or more prongs, an error which represents the "personal equation" of the age group. And the existence of an external copy which is the prototype of all the individual drawings might well be denied by a person who had to judge solely on the internal evidence afforded by the drawings themselves. But as the young draughtsmen gradually grow older the internal evidence undergoes a significant change. The range of variation gets narrower, and there is a tell-tale shifting of the average in one definite direction. The individual drawings get more and more alike; a larger and larger number of children put in exactly five prongs. And on this evidence alone it would appear highly probable that the wide disparities and variations in the children's earlier efforts were due, not to the absence of an objective standard, but to unfamiliarity with the objective

standard. The standard was there, but it was inadequately known. And its presence was inferable from the nucleating and stabilising influence which was perceptible in the children's efforts.

This is precisely the nature of the evidence we have for the objectivity of mental and moral values. Tangible prototypes, visible external scales, there are none. Nor indeed can there be, for the prototypes are essentially immaterial. But the internal evidence of their existence and objectivity is precisely the same as the internal evidence of the existence and objectivity of the original drawing of the five-pronged fork. There is the same tendency towards a consensus of opinion, a tendency which becomes more and more obvious as the individual minds develop. And to this can be traced what real authority there is in expert opinion, an authority which lies not so much in quantity of experience as in quality of experience.

It is admitted that our moral and æsthetic judgments vary perplexingly, that there are certain constant "errors" common to a coterie, a sect, or a nation; but it is contended that just as the divagations of the little children's drawings are due to immaturity, so are the divagations of our moral and æsthetic judgments due to our own immaturity and our failure to seize rightly the objective reality. In dealing with the simplest aspect of the physical world we are more or less grown-up, but in dealing with spiritual things we are still little children. And our blunders are little children's blunders, due to our fumbling after a dimly seen reality.

It is a mistake to believe that physical measurements differ from mental measurements in being

free from subjective taint. Indeed, all measurements in the last resort depend upon a guess or estimate by the man who measures. And however carefully he sets to work, and however precise the instruments with which he measures, there is always a margin of personal vagaries. Indeed, the greater the precision he strives to attain, the greater the likelihood that several successive measurements of the same thing will not be identical, but will vary round a central magnitude. It is only in crude measurements that identity of results is obtained, and this identity is due not to the attainment of absolute precision but to the presence of a constant error. As soon as high refinement is aimed at variations inevitably creep in.

Taking the whole realm of measurable things—things of which " more " or " less " can be predicated —physical and mental, material and spiritual—we see that it is possible (theoretically at least) to grade them in a series which begins with things in which the present possibility of objective measurement is at its maximum and ends with things in which the present possibility is at its minimum. And it should be the aim of all those who would advance human knowledge to press the possibilities towards the objective end of the scale—to extend as far as may be the application of exact measurement, and to eliminate as far as may be all those subjective factors that prevent us from seeing things as they really are.

CHAPTER V

THE REJECTION OF THE ESSAY

THE modern examination is dominated by the essay. It is based on the essay; it is built of the essay; it stands or falls by the measurability of the essay. And by the essay I do not mean a specially ambitious piece of writing; I mean anything and everything that may be called English composition—any "attempt," however trivial, to express ideas in discursive prose. Sometimes, as in the English paper, the whole test consists of one or two long essays and nothing else; more often, as in the History paper, it comprises from half a dozen to a dozen questions, to each of which the answer is a brief essay. Except in Mathematics and the Manual Arts the essay—the putting together of words—dominates the whole procedure. And the question here under discussion is not whether the writing of English is a profitable school practice (that, I take it, is beyond dispute); nor whether it is desirable that our pupils should go out into the world endowed with the capacity to express their thoughts in clear and correct English (nobody in his senses would deny that); but it is whether the essay, written under the stringent conditions of the examination room, and assessed in our present ignorance of the science of marking, is a good and true means of measuring. And my scepticism goes

further than denying that it is a good means of measuring the pupil's knowledge of geography, history, and science: I deny that it is a good means of measuring anything—even his knowledge of the mother tongue.

To doubt that an essay written at an examination, and written with a manifest desire to write well, affords trustworthy evidence of the writer's ability to write essays seems a degree of scepticism that amounts to perversity. And yet the facts justify the doubt. The first fact is that a pupil's essays show a greater variation in merit than any other intellectual product. The same pupil will, with equally good intentions, produce an excellent essay to-day and an atrocious one to-morrow. A well-known investigator, after much practice in appraising essays by the American method of comparing them with a standard scale of samples, estimates that the essays of one and the same student written within one year may cover a range of merit equivalent to six mental years. While one essay would scarcely match the achievement of a lad of twelve, another would do credit to a youth of eighteen.

That this seems to the teacher a gross exaggeration is due to the curious fact that in his marking of a series of essays by the same pupil he unconsciously tends to ignore variations and to mark them all on the same dead level. In other words, he does not mark so much on the merit of the piece of writing that lies before him as on his preconception of what the writer's ability actually is. The preconception may have been arrived at quite legitimately: it may, for instance, have been based upon what he regarded as the average achievement

of the pupil in his previous essays, or the pupil's average achievements in other lines of literary activity; but, however it has been arrived at, it is there, and strongly colours his view of what lies before his eyes. The consequence is that the slight variation in the teacher's scorings by no means indicates the real range between the pupil's high-water mark and his low-water mark. A story will make this clear. Many years ago—so long ago that only one of the three persons concerned is alive to tell the tale—there were two students at a training college. And they were friends. One was an Englishman named Smith and the other a Welshman named Jones. They were both members of an English class taken by the principal, and they had to send in every fortnight an essay on a prescribed topic. They worked together, they consulted together, they exchanged ideas, and they sent in pretty much the same sort of stuff. On the first occasion Smith's paper came back marked "Very Good," and Jones's marked "Very Fair." The same thing happened on the second occasion, and the third, and on nearly all subsequent occasions. If there was any change in the assessments it was such as would further emphasise the difference between the two essays. Smith's would be raised to "Very Good Indeed," and Jones's reduced to "Fair." One day they conspired together and laid a trap: they changed essays, and Smith sent in under his own name and in his own handwriting the essay composed by Jones, while Jones did the same with the essay composed by Smith. The expected happened. The paper signed with Smith's name came back marked "Very Good," and the

paper signed with Jones's marked "Very Fair." The principal was manifestly a good man, and just, according to his lights, and there was ample evidence that he read the essays with meticulous care; and yet it was clear that for some reason or other he had got it into his head that Jones could not write good English, while Smith could. Perhaps his previous intercourse with Welshmen had something to do with it. At any rate, there was the prejudice—the pre-judgment—which made it impossible for him to judge each essay objectively—to judge it by the witness it bore to its own worth.

This tendency of a tutor to mark his own pupil's essays in the light of a pooled impression blinds him to their wide variations, and, for good or ill, steadies the scoring far more than the several achievements warrant. The actual unevenness is much greater than he thinks it is. His misjudgment is at once detected when the essays are marked independently by a stranger. This should not surprise us if we remember the directive and steadying influence which a known name exerts on our judgment. The critic in *Fanny's First Play* hesitates to express an opinion of the play until he has learnt the name of the author. We are all like that. We are loath to believe that William Shakespeare ever wrote a bad line or that Ella Wheeler Wilcox ever wrote a good one. We look for satire and wit in everything that comes from Shaw's pen, and tenderness and charm in everything that comes from Barrie's. And what we look for in so complex a thing as a book or a play we are almost certain to find. Rarely do we behold a poem or a painting as a thing naked and

detached; we unwittingly invest it with something of the glory or the ignominy of the author's past.

The tendency for the merit of a pupil's essays to rise and fall (and especially to fall) is all the more pronounced if they are written under strict examination conditions—if a theme is sprung upon him on the examination day and he is asked to dilate upon it within the narrow confines of thirty minutes or an hour. During this brief time he has to mobilise and marshal his ideas (if he has any) and to set them forth in orderly discourse. Now the speed with which he can do this depends upon a multitude of factors, only some of which are relevant to the purpose of the test. The most important are his special interest in the prescribed subject, and the recency with which he has meditated upon it. When the topic is novel to him he has to start an organisation of ideas much of which takes place in the laboratory of the unconscious. And the slow organiser is no worse than the rapid organiser (probably he is better, as he has more forces to organise); but he always makes a worse show at the examination. The flow of ideas with every kind of candidate is always more or less fitful. And the difficulty of beginning is as great as the difficulty of coming to an end. It is like pouring out of a pot a viscous fluid that has formed an incrustation on the top.

The difficulty of beginning is always greatest in the self-critical; that is, in the more intelligent. Dissatisfaction with the second best impedes the progress of the essay, and the quantity produced is too meagre for the examiner to reach a just conclusion. As an instance of the disadvantage at

which the fastidious are placed let me cite the case of a little girl named K. B., who at nine years of age sat for a competitive examination at which she was required to write a story to illustrate the proverb: One good turn deserves another. Although her mental processes are extraordinarily rapid for her age, she is slow at penmanship and hesitant in phrase-making. The consequence was that the examination essay bore few signs of exceptional ability. When, however, she was asked the next day to write at leisure on the same topic this is what she produced:

A little stream meandered through a forest, the trees of which hung their branches over the brook. The wind rustled the foliage and seemed to whisper, "Hurry, little river, for thy father is lonely without thee," and, in answer, the current rushed on more swiftly than before. In the trees the birds sang sweetly.

Far more sweet, however, than the song of river, wind, and birds, was the song of a fair maiden as she filled her pail with clear water from the brook. Her golden hair was loose and floated in the breeze. She was clad in a long blue gown and her rosy feet were bare. After having filled her pail she rose and walked to a spot where stepping-stones formed a bridge. She tripped lightly from stone to stone, until, midway, she slipped and fell. The swift current would have borne her away had not a sturdy labourer's son arrived on the scene and pulled her out of the water, frightened, but fortunately unhurt. After thanking her rescuer repeatedly, she returned to her home.

Many years had passed. The fair Elaine had grown into a woman, and her youthful rescuer, a man. Still, in their memory the event of the riverside remained. His parents had died and he felt lonely. One day as he sat outside his little stone cottage which was situated on the outskirts of the forest, he mused upon his life of loneliness, and mourned for his dead parents. These sad thoughts so overcame him that he wept. As he was weeping Elaine approached him. She had been wandering through the woods until she came to the quaint little cottage hidden away among the cedar-trees. Elaine saw its owner's tears and she wondered what was the reason of them. She went towards him and gently asked why he wept. Encouraged by her kindly manner he told her his mournful story. When she had heard the tale, with tender words she soothed his restless spirit. When at last she left him to return home he felt as if he had been cured of that disease, sorrow, which weakens the soul and not the body. He had been cured by that wonderful talisman, Love. In that part of the country the birds sing as sweetly as of old, and the song they sing is called " Fair Elaine."

Is not this a remarkable piece of writing for a child of nine ? More remarkable and more significant still are the emendations which appear in her paper—emendations which give clear evidence of a critical and self-corrective attitude of mind that is extremely rare in one so young. Towards the close of the first paragraph, for instance, she first put " stream " instead of " current," and " quickly "

instead of "swiftly." In the second paragraph she deleted "voice" in favour of "song," deleted "pitc" (the beginning of "pitcher") in favour of "pail," and inserted the adjective "clear" between "with" and "water." Where she simply left "she slipped and fell," she originally had "she slipped upon a wet stone and fell half-way in the water and half on the large pebbles"; and so on all through. From beginning to end her paper is scored and scarred with deletions and corrections and interpolations. Sometimes her second thoughts are discarded in favour of her first thoughts. The last word but two, for instance, was omitted in the first draft; then "called" was wedged in; then it was crossed out and "entitled" put in its place; finally, this too was cancelled and the word "called" restored. An interesting change was made in the third paragraph, where she first put "she tried to comfort him," and then substituted "with tender words she soothed his restless spirit." The ordinary girl of nine is not embarrassed by a large store of words and phrases; she never has to choose between alternative ways of expressing a thing; she has but one way, and—if she thinks about it at all—thinks herself lucky to have even that.

Let us take another glance at K. B.'s mind. She writes poetry. This is what she wrote on the eve of her ninth birthday:

FAREWELL

Think of me, little sister,
When I am far away,
When the hills shall echo to welcome me,
Remember me night and day.

Sorrow you know is with me
And love within my soul,
Then pray to the Heavenly Father
That I may reach my goal.

Farewell, then, little sister
Think of me with thoughts as dear
As when the gentle Astrid
Followed her father's bier.

When asked about "the gentle Astrid" she replied that she saw the name in one of Longfellow's poems, and thought it a beautiful word.

Now here is manifestly a child with an extraordinary gift for letters; but she does not shine at examinations, and, indeed, it is doubtful whether any of the accepted types of examinations could ever reveal the full flower of her mind. But does she fare any better at the new examination? Most emphatically, yes! In the Silent Reading test to be found in pp. 147–51 of my book *Mental Tests*, the average score for 9 years of age is 6; for 14 years of age, 21; and for training college students, 41. K. B. was given this test, and she scored 47 marks. In fact she did better than 75 per cent. of the students at a training college. In my Chelsea Mental Tests, a group test of intelligence which gives a score of 45 for 14 years of age, she scored 68. She did all Burt's short series of Reasoning Tests. In my Standard Arithmetic Tests she showed the same degree of supernormality, the same superiority over the oldest of her school-fellows. This little girl has the intellect of an adult, and every type of new test to which she has been subjected (and she has been subjected to a large number) brings this fact clearly

to light. But the ordinary examination fails to do this; it fails even when it demands nothing but an essay; and her capacity for writing essays the reader can judge for himself.

We may lay it down as a general rule that an essay written at leisure and under the driving force of an interest which was either there to start with or was generated as the writer brooded over his subject, is a much safer index of ability than an essay produced under the pressure of examination. We may add that in either case it serves better as a positive than as a negative index: it is wiser to infer that the writer of a brilliant essay is himself brilliant, than to infer that the writer of a feeble essay is a fool.

It has been shown that one of the defects of the essay as a measuring device is the failure of any one example to represent the average literary ability of the writer. But there is another defect still more serious. It is the impossibility (the present impossibility at any rate) of making the essay amenable to rigid objective measurement. Even if any and every essay produced by the candidate, whether in the examination-room, the class-room, or the home, were a true and just sample of his literary powers, there would still remain the difficulties of marking—difficulties which are almost insuperable. Nobody can mark a given essay with a feeling of certainty that any other equally competent examiner would give the same mark, or, indeed, that he himself would give the same mark if he had to deal with the same paper again after he had forgotten his previous appraisal.

Some time ago I made the experiment of sub-

mitting seven essays to thirteen examiners. The essays, which were written by the boys in the top class of an elementary school on the subject, "An Ideal Holiday in May," were selected, not because they were nearly equal in merit, but because they presented different points of merit or defect. One was long, another short; one was florid, another matter of fact; one indicated a wide vocabulary, another a meagre vocabulary; one an observant habit of mind, another a meditative, and so forth. An independent examiner asked to mark them on a scale of 100, gave 40 marks to the lowest and 90 to the highest. So whatever the difficulties of marking may have been they were not due to small and indistinguishable differences in the essays themselves. The examiners, who were either head teachers or class masters who had had abundant practice in the marking of essays, were simply required to arrange the papers in order of merit. Now note the results. One of the essays occupied each of the seven possible positions, two of them occupied six positions, and each of the remaining four occupied five positions. No less than five-sevenths of the scale was covered by the essays upon which the agreement was closest.

The clue to this grave diversity of judgment is to be found in the complexity of the material judged. An essay is an intricate mental product which can be analysed in a variety of ways and yet can never be analysed completely. Each examiner consciously or unconsciously (as a rule unconsciously) makes his own analysis, measures each element by his own rod, weights each factor in accordance with his own scale of values, and finally

arrives at a verdict which sums up a number of hazy personal opinions. He calls it judging by general impression. But he is not very happy about it. The teacher who consciously marks his weekly tale of essays takes no great joy in the business. He would much rather mark sums. The sums are dull, uninteresting things in comparison, but he feels that he can mark them fairly. And knowing as he does that there is nothing that boys deprecate more in a teacher than injustice, or, as they prefer to call it, "favouritism," a vague fear of being unfair secretly assails him and helps to make the marking of essays the melancholy thing it is. For it is the marking or the criticising of the essays that is distasteful. The mere reading of them is pleasant enough. At any rate, there are compensations: in the desert of words a wonderful oasis may at any moment surprise and refresh the weary traveller.

But I need labour this point no more. Everybody knows and admits that an essay as an essay proves intractable stuff in the hands of an examiner; but few are ready to admit that every examination into which English composition enters is vitiated by the same taint as the essay itself. But it is. The man who marks an examination paper in Geography is willy-nilly marking English as well as Geography. He thinks he is estimating the candidate's knowledge of Geography, but all the while he is being unconsciously influenced by adventitious things—by the handwriting, the neatness, the spelling, the grammar, the comeliness of the wording—the thousand and one things that go to make up a series of written answers. It may perhaps be agreed that this is right and fitting; that knowledge

is so indissolubly wedded to words that you cannot test one without testing the other. The soundness of this plea may, with certain reservations, be admitted. It is not, however, the mere bringing of words into the business that is objected to: it is the writing out of the words in the form of an essay. This absorbs time and absorbs thought—a kind of thought that is totally distinct from the kind that is supposed to be tested. For the notion that we think in complete and well-rounded sentences—that valid thinking obeys the laws of syntax and of rhetoric—is founded on a false view of what goes on in our minds. The stupidest among us is always thinking about something (provided he is awake), thinking sluggishly perhaps and fitfully, but still thinking; yet no one would contend that the medley of words and phrases that flit through his mind would to an outsider make either rhyme or reason. If thinking were composing, then should we all be facile writers and fluent speakers. No, we have to piece together the rags with which our thoughts are clothed when they first enter our minds; we have to better the garments before we can show them to the world. The inner speech in which thought is carried on must be straightened out before the thought becomes communicable. And valuable as is this process of reshaping our inner speech into the conventional forms of written language, it is not History, nor Geography, nor Science: it is essay writing. And in examining we do not wish to count the same thing over and over again; especially if we cannot estimate it justly to begin with.

The orthodox examination in fact, with its dozen

questions to be answered by a dozen essays, is wasteful of time: it takes a prolonged sitting to test what may quite easily be tested in ten minutes. If from the two or three hours usually allotted to a public examination we take away the time taken up in putting ideas into decent English, and in the mere mechanical task of writing it out, there is not much time left for the amount of thinking that is strictly to the point. The astute teacher knows this; and when time is pressing and he really wants to find out how far his pupils have mastered a prescribed task, he distributes strips of paper and dictates a string of questions each of which is to be answered by a minimum of words. He realises that by this simple means he can find out more in five minutes than in a whole hour of formal examination. The pressure of circumstance has made him for the nonce a new examiner.

When a single reader marks all the papers at an examination, no question is raised as to the reliability of his marks. He is supposed to maintain a steady and uniform standard throughout. Yet it has been demonstrated by Dr. Starch that his marks are only about 1 per. cent. less variable than those of a panel of examiners working together on a preconcerted scheme. Great are the pains taken by such a panel to standardise the procedure and to make the marking objective. But, although we don't know it, objectivity is never reached. The examiner's blunders are as carefully hidden from the public gaze as are the doctor's blunders under the tombstones. But they sometimes leak out. The following instance is recorded in the Columbia University Report on the New Examination, which

is described in Chapter VII. At a certain American university an examination in History was held in the summer of 1920. The paper had to be marked by six professors of history working together as a panel. The failing line was 60 on a scale of 100, and all papers which a particular reader regarded as falling below this line had to be sent round to his colleagues so as to ensure justice to the borderline cases. One of the readers, who was troubled with an exacting conscience, wrote out for his own guidance what he considered a model set of answers. By some mischance this model paper got mixed up with his failing papers and was circulated among his colleagues for independent appraisal. It was read by them as a *bona fide* paper, and the marks it received ranged all the way from 40 to 80. The professor ran a risk of failing at his own examination.

Let it not be thought that in pointing out the weaknesses of the essay as a means of measuring I wish to speak lightly of it. I do not. My thesis is not that the essay is not good enough for a school examination, but rather that a school examination is not good enough for the essay.

CHAPTER VI

MICKLE AND MUCKLE[1]

THE new examiner pins his faith on the many mickles that make a muckle. In the field of intelligence testing, where the new examination methods were first developed, the use of a large number of small items is an essential part of the programme. Contrary to popular opinion, there is no such thing as a single test of intelligence. It is always one of a long series. The psychologist knows quite well that the wisest may easily be floored by a single test sprung upon him by a dolt. He knows that if a shaft from his bow fails to hit the target it is just as likely that his aim is bad as that the target is small. He makes allowances for stray shots. It is not his purpose to find out what is not there, but what is there. Of the 212 questions set at the American Army Intelligence Examination, anybody who successfully answered 135 or more was put in the first grade and labelled "a very superior person." So a recruit may have failed to answer over a third of the questions set and still have reached the highest grade. The critic of mental tests who picks out one test from a posse and confesses that it floors

[1] Etymologically the terms "mickle" and "muckle" are identical in meaning; but they are here used in the popular sense to which currency has been given by the proverb: "Many a nickle makes a muckle."

him should bear this fact in mind, and, if he is a modest man, console himself with the thought that he may fail in this, and in many another simple test, and yet have quite a respectable intellect. He may remember that even Homer nods. If, on the other hand, he is not a modest man, and his criticism implies that he, the testee, being a person of impeccable intelligence, failure on his part necessarily discredits the test, we can only advise him to guard his mind sedulously from tests of all kinds, and to pray for the fulfilment of the apocryphal part of this quotation :

> O wad some Pow'r the giftie gie us
> To see oursels as ithers see us !
> But how much better if by spells
> Ithers could see us as we see oursels.

Another misconception of a similar kind is that the new tests are merely " catches." As a matter of fact, the catch has no place in the new examination. While a test aims at measuring the mind, a catch aims at muddling the mind. If I ask an unsuspecting youth, " Which is the more correct, ' Five and seven *is* eleven,' or ' Five and seven *are* eleven ' ? " I am, under the pretence of asking him a grammatical question, deliberately leading him into an arithmetical trap. It is a catch, and not a test. A catch has a bias towards failure. A test is biassed towards neither success nor failure : it is quite neutral. If the testee elects to walk into a pitfall, that is his own concern ; but he may justly object to being inveigled there.

It is quite legitimate, however, to insert among a number of straightforward questions an occasional

question designed to distinguish between the circumspect mind and the routine thinker. My arithmetic test, for instance, includes the following question :

" A man takes 20 minutes to walk from his house to the station. His son also takes 20 minutes. How long will it take them if they both walk together ? "

This may, perhaps, be described as a " catch " ; but it is a legitimate catch, as it does not stand by itself as an isolated test, but is a mere element in a multiple test ; and it can catch none but the stupid and the thoughtless.

The main advantage of a large number of questions is that they cover a wide field. Instead of probing the mind at a dozen points they probe it at a hundred points. The old examiner dips his hand a few times into the storehouse of the pupil's mind and brings up samples which he assumes to be fairly representative of the whole stock. But chance has a hand in it as well ; and the less frequently and less widely the examiner dips, the more active is the hand of chance. When in a particular instance the examiner has taken a bad sample the candidate explains to his companions that he has had " rotten luck " ; to his master, that he has not had an opportunity of doing himself justice. The less " luck " there is the better ; and the more questions are squeezed in the more luck is squeezed out. Thus the new examination gives the candidate a greater feeling of satisfaction than the old ; a feeling that at any rate he has got his deserts—that the examiner has put his finger on his strong points as well as his weak points.

It is no small virtue of the new examination that it defeats the crammer. The term " cramming "

seems to be used in two distinct senses. It sometimes means the stuffing of facts into the mind in such a way as to render digestion and assimilation impossible. As the end and aim of the crammer is to get his pupil through an examination, and not necessarily to give him a mastery of the subject, it does not matter to him that the pupil's grasp of the material is feeble and uncertain. If it is held in any fashion till the examination day, it may, for all he cares, be dropped the day after. But cramming, or special coaching, may take another form—the form of confining the study of the subject to those parts that " pay " at an examination. It involves a study of the mind and methods of the examiner, and a close scrutiny of past questions, so that the actual questions set at the examination may as far as possible be anticipated. It means a crowding of the goods at the places where the examiner is likely to place his hand—a deliberate attempt to falsify the sampling.

In the former sense of the word, cramming for the new examination is manifestly impossible. To cram for the intelligence test is labour in vain, for the test is not concerned with facts but with the mental manipulation of facts; a deal of hard thinking has to be done on the spot, and often in a limited time. The capacity to deliver the dead product of past thinking (especially if it is somebody else's thinking) is of no use at an intelligence examination. Nor is it of much use at an attainments examination—not if the questions are designed, as they should be, to test the pupil's power to think in the special material, and not merely to remember the special material.

As for the second sense of the word " cramming," the danger of the examination questions being known, either by accident or by cunning anticipation, is far less serious when there are a hundred questions than when there are only ten. To anticipate the scanty questions of an ordinary examination is a distinct advantage to the candidate; it limits his field of study and saves his time. But to anticipate the multitudinous questions of a new examination will profit him nothing; for they cover the whole course of study. The only way to cram for the Algebra examination, given on pp. 200–202, is to work steadily through every exercise in an elementary textbook; and when this is done the word " cramming " loses its meaning. There is no distinction between it and hard, honest work.

Complaint is sometimes made that children are coached for the Binet tests. Why they are coached is not known. In any case, the alert examiner has little difficulty in detecting the coached pupil; for the coached pupil has a special aptitude for " giving himself away." And for the suspects the examiner always has a few alternative tests up his sleeve. Group tests of intelligence are equally cram-proof. Dr. R. R. Rusk, who devised the excellent series of group tests used by the Fife Education Authority, has investigated the effects of coaching on this type of test. He maintains that " the effect of coaching can be discounted by the possession on the part of the examiner of a wide repertoire of tests. By a slight modification of the form of recognised tests the factor of negative transfer may come into play, and pupils who are coached may find themselves at a disadvantage in

attacking the modified form of test." [1] He goes on to show that in schools where coaching was reported to have taken place the children did a little worse than in schools of similar standing where it was certain that no coaching had been done. In fact, an attempt to coach for an intelligence test defeats its own end.

It has been wisely said that no examiner should draw up an examination paper without asking himself the question: How is this going to affect the teaching? It is idle to urge that the class teacher should ignore the pressure of examinations imposed upon him from without. He cannot ignore it. If his pupils have to be examined at all he will inevitably "play up" to the man who examines them. If the examination of the future is to consist in the main of a large number of little tests, he will practise his pupils in performing a large number of little tasks. Well, what of that? The little task is better than the big and complex task, not merely as a means of testing, but also as a means of training. The change from large to small, if it ever does take place, will be comparable to the change that has already taken place in the realm of physical culture. There was a time when it was held that the proper way to develop the muscles of the arms and chest was to use large and heavy dumb-bells. Now the dumb-bells used are small and light. It is said that the late Colonel Burnaby used to practise with dumb-bells weighing seventy pounds. It was then thought that the heavy dumb-bells were the cause of his strength;

[1] The *Scottish Educational Journal* for December 15, 1922, p. 895.

it is now thought that they were merely a proof of his strength. The modern boxer is trained by a number of brief bouts frequently repeated; the modern runner is trained by a number of short runs at regulated intervals; the modern gymnast is trained with light exercises and light apparatus. The old athlete was notoriously short-lived. In those days no distinction was made between physical training and physical straining: it was not realised that big biceps were often got at the expense of a weak heart.

Many a teacher has already discovered for himself the value of the little task and the little test. In my own teaching days it gradually forced itself on my notice and justified itself in my practice. I was at one time in charge of a chemical laboratory and taught chemistry to pupil teachers. Well do I remember the care and trouble with which I once got up a lesson to show how hydrogen may be prepared by passing steam over red-hot iron filings. I spent a whole afternoon (a holiday afternoon, too) in routing out and furbishing a Becker's Furnace which had been out of use for years, and fixing it up for the next day's lesson. The lesson itself was, to all appearances, a huge success. The demonstration table presented a fine display of Bunsen Burners and tripod-stands and flasks and gas-jars and hydrogen balloons and the various paraphernalia which a painter loves to put in an alchemist's den. And the experiments came off without a hitch. The students would no doubt have liked it better if there had been an explosion, or a general smash-up; but I myself was quite pleased. I was anything but pleased, however, when

I tested my pupils later on; for I found that though they had remembered many of the details of the lesson they had failed to grasp the essential points. They were great at description—of the wrong things—and many of them called the Becker's Furnace a Beggar's Furnace; but few were able to say where the hydrogen came from. Ever after I aimed at simplification. Irrelevant things were hidden behind a screen, and each experiment was stripped bare of all accessories. And the students had to make and to repeat numerous little experiments in the laboratory rather than spend hours on fixing up elaborate apparatus. It took but a minute to make hydrogen in a test-tube by dropping zinc into acidulated water; and the test-tube was as good as a Woulffe's bottle for getting familiar with the essential materials and the essential process. It was by simple little experiments that I was able to secure a reiteration of the important.

So, too, in mathematics. A reiteration of the important I found possible by contenting myself with little sums. The little exercise and the little problem, each with a point of its own, enabled me to discover my pupils' weaknesses, and to test at frequent intervals if those weaknesses were being overcome. It was by the little exercise and the little problem that I was able to cover a large part of the course in a short time. As the years rolled on and new relays of students passed through my hands I got more and more convinced that for the triple purpose of diagnosis, training, and examination the small example, so long as it embodied a point or a principle, and challenged the pupil's thought, was of far greater value than the larger

and more complicated examples of the textbook. For purposes of testing it is important that each little problem should have a point—a fresh point —a point of its own—so that it is impossible to solve it by the mere momentum of thought set up by the previous problem. It seems to me that on this ground some of the standard tests in use in America are open to criticism. For example, Test 4 of Rugg and Clarke's Standard Algebra Tests requires the solution of the following equations among 25 others; (6) $3x - 4 = 16$, (11) $5x - 2 = 27$, (16) $4x - 5 = 17$, (21) $6x - 3 = 33$. These equations are of precisely the same form and test precisely the same algebraic process. The differences are arithmetical not algebraic: and it is not arithmetic that is being tested.

Many teachers attach enormous importance to mental arithmetic: they invest it with sovran virtues. They claim that it brightens up the children's minds, sharpens their wits, and kindles their intelligence. If, indeed, all they say about it is true, it is difficult to see what need there is for including anything else in the curriculum. But ignoring the formal training fallacy that lurks in this piece of special pleading we can readily admit the merits of mental arithmetic, especially when its advocates point out that with oral exercises much more real arithmetic may be got through in a given time than with written exercises. Mr. Marshall Jackman has abundantly demonstrated in practice that much time is saved, and no efficiency is lost, if written arithmetic is postponed till the children reach Standard III. There is not the slightest doubt that mental arithmetic signally succeeds.

It succeeds, not because it is mental (all arithmetic is mental), but because it involves a multitude of little tasks, each standing for the while in isolation, attacked by itself, assessed by itself, criticised by itself. It is searching and it is rapid. It probes the crevices and corners and reveals the weak spots. These are its virtues. They are the virtues common to the little task and the little test. They do not, as is generally believed, consist in a general strengthening of the mind by the carrying out of long and involved calculations without putting pencil to paper. The ability to work sums in one's head is largely dependent on a sort of numerical imagination—an innate capacity for steadily holding numbers before the mind's eye—a capacity which can be shown to co-exist with inferior mathematical powers. It is doubtful whether such a faculty can be cultivated at all; or, if it can, whether it is worth cultivating. The reader has, no doubt, met the kind of man who has an amazing faculty for figures—for making complicated calculations in his head—but is such an egregious ass in other respects that one is not encouraged to emulate his mathematical feats.

Whenever in the past I have put forward a plea for the brief example in arithmetic I have generally evoked the objection that the brief example gives no training in "sustained effort." The reply is that there is no such thing as a general faculty of "sustained effort"; and that effort, as the Herbartians have convincingly taught us, is more a matter of interest than of training. For effort, far from being opposed to interest, is indissolubly yoked with it, and cannot be put forth without it.

Effort is, in fact, generated not by previous effort, but by the supply of motivation. If interest in the task is there effort will be there too. Whether it is so or not, it remains as a demonstrable truth that the more excellent way to prepare for big sums is by working many little sums rather than by working the big sums themselves. It is the question of light dumb-bells versus heavy dumb-bells.

Having said so much for little things let me remind the reader that the tasks we have to perform in daily life are neither all big nor all small; they are of assorted sizes. So should be the exercises in school; and the tests. Big things should occasionally be done. But for the purpose of gaining power through practice, for the purpose of revealing the pupil's points of weakness and of strength, and for the purpose of measuring ability and attainments with a reasonable degree of exactitude, there is nothing to compare with the small task and the small test.

CHAPTER VII

THE NEW EXAMINATION ON TRIAL

In recent attempts to test the intelligence of a large number of people at once, psychologists have been forced to develop the technique of examination and to make new and tentative departures. They have used devices never used before. And some of these devices have proved so successful in the testing of intelligence that they are now being tried for the testing of attainments. They are passing from the hands of the psychologist to the hands of the teacher. But the technique is sometimes so startlingly novel that many a teacher looks upon it with grave suspicion. While claiming to be more precise it appears at first blush to encourage guessing, and to lead to conclusions which seem likely to prove even more uncertain than those reached by traditional methods. But when put to the only test that is of real value, the test of scientific experiment, the new technique emerges triumphant. It satisfies better than the old technique all reasonable criteria. "It works." It works as well for attainments as for intelligence, as well for an entirely fresh examination as for an examination already standardised, as well for the casual and informal examinations held in the class-room as for the formal examination with its printed paper and official invigilators. It is not impossible indeed

that the new devices will ultimately prove even more useful to the teacher than to the professional examiner. The psychologist has, at any rate, put into the teacher's hands new educational instruments which he would be unwise to reject before testing their serviceableness.

An important investigation into the value of the new technique was made at Columbia College, New York, in 1921, and a report, prepared by Mr. Ben D. Wood, assistant to the Dean of the College, was published in the September issue of *Educational Administration and Supervision*. The subject selected for the experiment was "Contemporary Civilisation." Turning a deaf ear to the whisper of the cynic that the subject-matter does not exist, let us hear what the report has to say. It claims that "the Course in Contemporary Civilisation presents ideal conditions for a thorough-going trial of various kinds of measuring devices. This course, since it is required of all freshmen, includes a large number of students of varying ability, studying the same material as indicated in a very detailed syllabus, in the same time, by the same method of instruction and recitation, and under the direction of an unusually harmonious staff of instructors. In addition there was available a reliable objective mental classification of the students of the course, as well as other data indicative of what might be expected of them." The term "mental classification" refers to the findings of the Thorndike College Entrance Intelligence Examination, which all candidates for admission were required to take.

The experimental examination which was set at the end of the course was taken by 440 freshmen.

It consisted of two parts: the first, which lasted an hour, representing the old examination; and the second, which lasted two hours, representing the new. The old examination required the answering in essay form of two questions out of four. I quote as an example one of the questions:

Define the major groups or classes in England or in France in the early eighteenth century, and show in some details the conflicts of these groups and the changes wrought in them and their relative position by the political, economic, and social changes of the eighteenth and first half of the nineteenth centuries.

The new examination comprised three tests, the first of which was of the True-False type. It consisted of 139 statements, each of which had to be marked with a plus sign if the examinee thought it true, and by a minus sign if he thought it false. I will cite a few of the statements:

A.9. The civilisation of the Phœnicians, Greeks, and Romans centred about short river valleys.

12. Irrigation is carried on more extensively on the eastern than on the western slopes of the mountains of California.

B.16. Among animals under conditions of domestication instinctive behaviour tends largely to be displaced by habitual or acquired modes of behaviour.

18. The enclosing of the "Commons" brought about a great increase in the number of small holdings.

51. By his conquests in Central Europe Napoleon retarded German unification by at least a hundred years.

The score for this test was the number right minus the number wrong.

The second test consisted of 14 paragraphs, in which certain missing words were to be supplied. This is technically known as a Completion Test. The eighth paragraph is here quoted:

8. By the mercantilist policy the European Powers sought to increase their . . . and to decrease their . . . Especially the . . . of raw materials and the . . . of finished products was encouraged. A representative mercantilistic statesman was . . .

The third and last test was a Recognition Test. Here 61 statements were presented to the candidate, who had to underline the one word or phrase which would make each statement true. It involved a choice of several possibilities. I quote three items :

A.5. Most of the inhabitants of the British Isles belong to the race known as the—Alpine, British, Teutonic, Mediterranean, Basque, Celtic.

B.7. Miserliness or kleptomania are abnormal exaggerations of the instinct of—Hunger, Love, Acquisitiveness, Fear.

C.30. The passage of the Reform Bill of 1832 was a victory for the—Agricultural Landlords, Farm Labourers, Factory Workers, Well-to-do Middle Class.

Here we have a group of tests carrying altogether 250 points of credit, tests which satisfy all the essential conditions of the new examination. The items are small and numerous, they involve a minimum of writing, they are fool-proof, and they cover the whole course of study. The " question paper " was not a paper but a booklet. The trouble of preparing such a booklet was compensated for by the ease of marking. The preparation, however, was not very onerous, for the instructors in the course of Contemporary Civilisation (and there seems to have been at least a dozen of them) formed a committee, which divided the work amongst its members.

A cursory glance at this example of the new examination might give one the impression that it places too much emphasis on the memory of

isolated facts; but a closer scrutiny shows it to be a severe test of judgment and of the power to organise the material studied.

Now comes the important point: how were the authorities to judge the value of the new examination? The obvious way was to compare its verdict with other criteria—other estimates of the same thing. The criteria actually used were threefold: the Thorndike Intelligence scores, the instructor's estimate of achievement (based on class work), and the result of the one-hour examination. There was in addition the evidence afforded by the distribution of marks. All the available data indicated that the real abilities and attainments of the students were distributed normally—that they roughly accorded with the probability curve—and it is significant that the scores of the new examination fitted this curve far more closely than did the scores of the essay examination. The initial presumption, based on statistics, pointed to the new examination as the better of the two.

The most cogent evidence, however, in favour of the new examination was furnished by the correlations between the various estimates referred to above. It has been found by extensive research that the more carefully and accurately school marks are accorded the more highly do they correlate with general intelligence, but only up to a certain limit—up to ·5 or ·6. Now the coefficient of correlation between the new examination and the Thorndike intelligence scores was ·513, and this was a higher index than that given by either the essay examination (which was ·386) or the instructor's estimate of achievement (which was ·358).

This in itself proclaimed with much convincingness that of the three ways of measuring attainments—the tutor's opinion, the essay examination, and the new examination—the new examination was considerably the best.

The sceptic may urge that the new examination correlates highly with the intelligence examination simply because both test the same sort of thing. If that were so it would have lower correlations with the essay examination and with the tutor's estimate. In point of fact, they were not lower; they were higher; they were ·654 with the essay and ·620 with the tutor's estimate. The unavoidable inference is that the new examination really measures what it purports to measure—the students' attainments.

Now mark this new line of argument. It is reasonable to assume that where there are several guesses at a magnitude the average of the guesses is nearer the truth than any one of the separate guesses.[1] The principle does not always hold good because of the occasional presence of a constant error—a drifting of all judgments in the same false direction; but in the main the principle is sound: the average of diverse estimates is the most likely to be the true estimate. In the investigation which we are considering there were two independent guesses—the instructor's estimate, and the scores of the essay examination—and there was no reason to suspect a constant error. The average of these two guesses was assumed to be nearer the truth than either of

[1] Professor Edgeworth lays this down (in different and more precise terms) as one of the postulates of examining. See Hartog, *Examinations*, p. 103.

them taken by itself. And it was found that the new examination correlated more highly—tallied more closely—with the average of the guesses than with the individual guesses. And this was true not only of the new examination as a whole, but also of each of the three tests taken separately. The more trustworthy the standard with which the new examination was compared the better did it appear.

The same line of argument was applied again: to the essay examination exclusively this time. A number of the essay booklets (117 altogether) were chosen at random and marked afresh by different examiners. Thus were secured two independent appraisals of each of 117 booklets. After saying that the correlation between the two sets of marks was no higher than ·663 (I make no apology for calling the marks guesses. Here again then we have two guesses—the first marking of the essays and the second marking of the essays. The correlation of the new examination with the first guess was ·599; with the second, ·533; and with the average of the two, ·623. Once more the new examination showed up better when the standard of comparison improved.

When it comes to the question of reliability, the new examination wins easily. By the reliability of an examination is meant the steadiness with which it measures. If it is given the second time to the same candidates and the results are nearly, if not quite, the same as those originally obtained, it is technically described as reliable. It would be unreasonable to expect identity in the results, for the thing measured has itself undergone a change;

but the correlation between the two sets of results should be fairly high. It is often, however, inexpedient, and sometimes impossible, to give the same examination twice over. Fortunately we have another way of testing its reliability. If the questions are numerous enough we can pretend that there are two examinations instead of one; we can pretend that all the questions with even ordinals (second, fourth, sixth, etc.) constitute one examination and that all the questions with odd ordinals constitute another examination; and we can compare the two sets of marks. (We can compare them informally by inspection, or, better still, we can compare them with mathematical precision by the method of correlation.) If the two sets are alike, the reliability is high; if not, it is low. The new examination was treated by some such method as this, and the coefficient of reliability was found to be ·905.) This is exceptionally high; and although it was not possible to calculate the reliability of the essay examination in the same way, there was every reason to believe that the reliability was very much lower. If the correlations between two markings of the same set of essays was only ·663, there was no ground for believing that it would have been higher if the essays themselves had been changed as well.

On the score of objectivity there was no comparison between the two examinations. While the objectivity of the new examination was perfect, and the correlation between the marks of two independent examiners of the same booklets was 1, the correlation between the two independent markings of the essays was no higher than ·663; and this

in spite of the fact that the essays in question were marked, as the report assures us, with much greater care than is usual at an examination.

Mr. Wood's report ends with testimonies given by some of the instructors who sat in judgment on the new examination. They were at first sceptical as to its value, but the experiment converted them. Since the testimonies are all of the same tenor, I will quote from the first only, that of Mr. Austin P. Evans, Assistant Professor of History. He says :

> All the evidence seems to point to the conclusion that it (the new examination) is much the most accurate method of testing students' knowledge of the subject that we have yet discovered. It does not, however, in my judgment, test sufficiently the student's powers of accurate and cogent expression, and his ability to organise his material. I should therefore feel that in an examination adequate time should always be given for at least one or two questions of the essay type.

The others agree with Mr. Evans in praising the new examination, and in pleading for the retention of the essay as part, but part only, of future examinations.

CHAPTER VIII

THE TRUE-FALSE TEST

THE true-false test so often used in gauging intelligence is signally open to attack. Consisting as it does of a number of statements which the examinee is asked to label true or false, it seems a wanton incitement to the tossing of a coin. Whatever he says is just as likely to be right as wrong. So when he does not know, he chances it and says anything. If he calls all the items true he will get about half of them right; if he calls them all false he will get about half of them right; if he calls them true or false at random he will get about half of them right. So the die is always loaded in his favour: he has nothing to lose by guessing and much to gain. Or at least it would be so but for the marker, who has a scheme for spoiling the guesser's game: he neutralises the effect of guessing by taking away a mark for every item that is wrongly labelled. Suppose, for instance, there are 60 statements, and the candidate has no knowledge of the subject-matter but guesses all his answers. He will, on the most probable estimate, get 30 items right and 30 wrong; and his score will be $30 - 30 = 0$. Suppose, again, he deals with 50 of the items by actual and accurate knowledge, and with the remaining 10 by chance. About 5 of his chance shots will hit the mark, and

his final score will be $55 - 5 = 50$. Thus he gets credit for his knowledge and none for his guesswork.

The gambler in true-false marks is in fact hoist with his own petard. The very laws in which he trusts—the laws of chance—are used against him. The scheme of scoring ensures that as a general rule, and in the long run, his losses will cancel his gains, and he is left in the same position as at first. As a general rule and in the long run; for it is a characteristic of the laws of chance that they gain in certainty as they lose in particularity. They are truer of 100 cases than of 50 and truer of 50 than of 10. As the number of random judgments in the true-false test increases so does the number of right judgments tend to bear to the number of wrong judgments the ratio of one to one. So the more the guesser guesses the less has he to gain by guessing; and the less has he to lose. It is in occasional and casual guessing that the uncertainty is greatest. But the uncertainty works both ways: it is just as likely to work against the guesser as in his favour; just as likely to bring him unmerited blame as to bring him unmerited credit. If, for instance, in the supposititious case above, where the candidate judges partly by knowledge and partly by guesswork, he guesses 6 correctly out of 10 (not an unlikely contingency), his score will be $56 - 4 = 52$, and he will have gained 2 marks beyond his deserts. But, on the other hand, he is quite as likely to guess 4 right as 6, and his score will then be $54 - 6 = 48$. Here he will have got 2 marks short of his deserts.

Sound, however, as the true-false examination seems to be as a whole, it still remains as an undeni-

able blemish that if each individual item is taken by itself, the candidate's choice offers an inadequate assurance of real knowledge. He has hit one mark out of two, instead of one out of many; so we cannot be sure that his aim was deliberate. But this only discredits the diagnostic value of the test; it does not discredit its mensural value. Its mensural value steadily improves as the number of items increases. And when 50 or 100 items are used there are reasons for believing that it affords a more accurate measure of knowledge than the traditional type of examination.

There are two distinct ways of presenting the true-false test, one which invites guessing and another which prohibits it. When the test is used for testing intelligence, as in the American Army Examination, the instructions to the candidate always say, "If you cannot be sure, guess." When, however, it is used for measuring acquired knowledge, the candidate is sometimes warned against guessing by being shown its futility. In the Columbia College experiment described in the last chapter the directions for the true-form test are so good that I quote them in full:

Read these statements and mark each one at the left of its number with a plus sign if you think it is true, with a minus sign if you think it is false. Each statement marked correctly gives you a credit of one point; each incorrectly marked statement counts as a penalty against you, and is subtracted from your score; omitted statements count neither for nor against you. Your score will be based on plus and minus signs; don't waste time writing anything else. First, go through the list quickly and mark all that you know for certain at once; then go back and study out the harder ones. Do not guess. The chances are against you on guessing. A wrong response counts heavily against

your score. Don't endanger your score by gambling on those questions about which you know nothing.[1]

One reason for asking the candidate to guess in the true-false test at an intelligence examination is that a rigid time limit is imposed. The statements themselves are ridiculously easy; as, for instance, "Gunpowder is not good to eat," or "Butter and cheese are made from eggs." There is no difficulty whatever in saying whether the statement is true or false; the only difficulty is in saying it quickly. It is not a test of knowledge but of alertness of mind—quickness of apprehension—and it is important to show the number of meanings grasped in the given time. This is not shown if any items are read and then skipped. Another reason is facility of scoring. There are two formulæ for scoring. One is: $R - W$, where R means the number of items right and W the number of items wrong. The other is: $T - 2W$, where T is the total number of items attempted. These two formulæ are identical; for, since $T = R + W$, if we take $2W$ from each side we get $T - 2W = R - W$. The formula $T - 2W$ is much the easier to apply—especially if the candidate has worked steadily down the list of items and made no omissions.

On the whole, it seems reasonable to discourage guessing in the true-false test, especially when it is used as a test of attainments as distinct from a test of intelligence.

Mr. F. B. Knight, of the University of Iowa, who

[1] See *Journal of Educational Psychology* for February 1922, pp. 75–80 (vol. xii, no. 2).

examines his students in Physics by the true-false method, presents the test thus[1] :

If statement is true put a T in column headed T.
If statement is false put an f in column headed f.
If you are uncertain put a U in column headed U.

	T	f	U
1. Light must fall upon the objects themselves if we are to see them			
2. In order that a body may be seen light must pass from it to the eye, and usually this takes place along straight lines			
3. No appreciable time is required for light to pass from one point to another			
4. The velocity of light in water is less than in air			
5. Electric waves have the same velocity as light			
Etc.			

Not only does this arrangement encourage candour, but it also makes the marking easier; for the answers are doubly distinguished—by initial and by position.

Mr. P. L. Gray and Mr. R. E. Marsden, in an experiment which they recently made with the Terman Group Test of Mental Ability, paid much attention to the true-false questions included in the test. They tried to discover by experimental means to what extent children actually guessed instead of answering from knowledge; and they came to the conclusion that the amount of guessing was so small as to be almost negligible.[1] So convinced, indeed, are some examiners that guessing does not take place in the true-false test, especially when three possibilities are presented instead of

[1] See *Educational Research*, Supplement to *The Head Teachers' Review*, for May 1922.

two, that in the scoring they ignore the wrong answers altogether and simply count the right ones.

We have now to discuss the validity of the true-false test. Can it be safely used as a substitute for the ordinary examination, formal or informal? Even if it were not a whit more trustworthy than the essay examination, it has so many points of merit, so many practical advantages, that its use—its occasional use at any rate—would prove a distinct boon to the teacher. But there are good grounds for believing that it *is* more trustworthy than the essay examination. All the evidence brought forward in the last chapter to support the claims of the new examination supports equally the claims of the true-false test. For nearly half the questions at that examination were of the true-false type. Moreover, all the correlations to which I referred had been calculated for the three tests separately as well as for the examination as a whole; and the coefficients for the three tests were in every instance so nearly equal that the differences between them were negligible. The true-false test proved to be just as valid and just as valuable as the completion test and the recognition test.

A more extensive, if not more thorough, investigation of the worth of the true-false test has been made by Mr. Arthur I. Gates of Teachers College, Columbia University, who used it at various times to examine ten separate classes in educational psychology, each class having about 70 students.[1] The standards with which he compared the true-false test were an essay examination, written work

[1] *Journal of Educational Psychology* for May 1921 (vol. xii, no. 5), pp. 276–87.

(at home and in class), an intelligence examination, and what Mr. Gates calls "the Criterion." "In constructing the criterion, the sum of all essay examinations was given the weight of 1·0, the sum of the true-false tests, 1·0; the sum of the written work, 0·5; the sum of the class recitations, oral quizzes, special conferences, etc., 0·5."

A striking characteristic of Mr. Gates's work is the frequency with which he confirms his statistical findings. Instead of being content with calculating once only the correlation of one true-false test with another true-false test, he did it over again with another pair, and yet again with another pair, and so on until he had got 59 correlations. Then he took an average. The individual correlations ranged from ·13 to ·81, and the average was ·54, with a standard deviation of ·152. This was a higher result than was obtained by the self-correlations of any of the other modes of examining. For the correlation of one essay examination with another was ·36, and of one measure of written work with another ·43. It was found that the correlation of the true-false test with the essay examination was as high as the correlation of one essay examination with another essay examination. That is to say, a true-false test affords just as sound a basis for inferring the score of a student's next essay examination as does his last essay examination.

When the several methods of estimating the students' achievements were compared with the Criterion, it was again found that the true-false test came off best. The correlations were ·65 for the true-false test, ·56 for the essay, and ·39 for the written work. More significant still was the

marked rise in the true-false correlations when two or more results were compounded. For as the number of true-false examinations taken together increased from 2 to 5, the correlations with the Criterion became respectively ·672, ·701, ·744, and ·803. When the essay examinations were similarly pooled the corresponding correlations were ·541, ·556, ·573, and ·610. Hence, the gain with pooling is very much less. Five essay examinations scarcely afford a more reliable index than one essay examination, whereas in the true-false test the quality signally improves with the quantity.

The same trend was revealed when the various estimates were compared with the results of the intelligence examination. A single true-false test correlated with intelligence more highly than any other single test. The co-efficient was ·406, as compared with ·344 for the essay and ·255 for written work. And when several results were taken together the superiority of the true-false test became still more apparent. The correlation of two true-false tests with intelligence was ·432, of three ·472, of four ·508, and of five ·545. Contrast these with the correlations of the essay examinations with intelligence. They were ·340 when two were grouped, ·344 when three were grouped, and ·350 when four were grouped.

We are not surprised to find this pronounced improvement in the validity of the true-false test as the number of tests increases. It accords with what we perceived at the outset: though much uncertainty rests in a single true-false question, that uncertainty is gradually eliminated as we add more and more questions. If the number of items in

each of Mr. Gates's tests had been larger (he says that each "consisted of a series of 30 or more statements") the correlations of the single test would no doubt have been higher. They would have agreed more closely with those recorded by Mr. Wood. Pooling the separate tests had precisely the same effect as increasing the number of items in a single test. They acquired respectability as their numbers multiplied.

One advantage of the true-false test is that it can be given in more ways than one—that the mode of procedure may be varied to fit the amount of time at the disposal of the examiner. He may laboriously print off copies for the whole class, or he may write the test on the black-board, or, if he is in a hurry, he may dictate the test item by item.

But the outstanding merit of the true-false test is that it is didactic—that it teaches as well as tests. It may be objected that all tests teach. And so they do in a sense, but the true-false test is pre-eminently a teaching test: it has a peculiar potency in building up systems of knowledge in the pupil's mind. I assume, of course, that the method of marking described in Chapter XII is adopted—I assume that the test is corrected immediately after it has been given, and that it is corrected by the pupil himself while the teacher reads out each item and announces its truth or its falsity. If this is not done, if the pupil never sees his paper again, or only sees it after a lapse of time, or is not forced to rethink his former thoughts and to confirm or reform them in the light of his teacher's ruling, the didactic value of the test disappears.

It is no longer a mentor, but only a measure. If, on the other hand, the rules for constructing and marking are rigidly observed, it will be found, I think, that the children will learn more from the true-false test than from any other type of examination. And it is a curious fact that they will learn more from the false items than from the true.

The reader will ask, and rightly ask, for evidence. Here it is. The geography examination which appears on pp. 210–214 of this book contains a true-false test with 45 items, 22 of which require "yes" for an answer, and 23 require "no." The test was given to a class of boys about 13 years of age. Each marked a companion's paper, and afterwards checked his own. The average mark for the true items was 12, and for the false items 11·75. All the test papers and answer papers were collected, fresh papers were given out, and the children were again tested in the same subject-matter, but in quite a different way. Each of the 45 items was dictated in the form of a question requiring a simple definite answer which would reveal whether the pupil had really mastered the point at issue. For example, the items numbered 61 and 62 read as follows:

61. Southampton is the second largest port in England.
62. York is an old Roman town.

During the correction something of this kind was said to the class:

"'Southampton is the second largest port in England.' No, Liverpool is: it comes next to London."

"'York is an old Roman town.' Yes!"

In the re-testing these items were dictated in this form :

61. Which is the second largest port in England ?
62. What nation first built the town of York ?

As a result the average mark for the items which originally had been true was 14·1, and for the items which originally had been false was 15·44. There had therefore been an improvement of 17 per cent. on the true items and 30 per cent. on the false.

The same type of experiment was made with the history examination (see pp. 224–229), with still more surprising results. There was in one school an improvement of 7 per cent. on the true items, and of 73 per cent. on the false ; in another an improvement of 14 per cent. on the true and of 64 per cent. on the false. In no case were the children told that they would be retested ; nor were they urged to remember the facts on which they had been examined. They simply marked the papers with no anticipation of further inquiry. It almost seems as though we could teach children better by first telling them lies and then telling them the truth than by telling them the truth twice over.

The statistical evidence of the didactic value of the false items came to me as a surprise, for my mind had been imbued with the warning that used to appear in all the old manuals of method. Never present anything to the class in an incorrect form. It mainly referred to spelling : it aimed at discouraging the practice of putting on the blackboard mis-spelled words for the children to correct. And there is no doubt that in the matter of spelling, and,

indeed, in the whole realm of things that have to be learnt purely by imitation, or by rote memory, with no rational explanation of their rightness or their wrongness, there is wisdom in the warning. Unless a child is given to argue about spelling (and he is rarely interested enough in spelling to argue about it) there is no point in telling him that "c-o-f-f" does not spell "cough"; for he is not in the least likely to encounter the word in that form. The children who were warned by a fussy grandmother not to push buttons up their noses were rather glad to get the suggestion. It was a form of sport which they had never thought of.

In the ordinary intercourse of life, however, we cannot escape the *suggestio falsi*; we cannot avoid getting our minds muddled with truths masquerading as falsehoods and falsehoods masquerading as truths. If we critically examine our ideas on any topic we shall find strands of truth and error confusedly intertwined. It is difficult to say which is which. And clear thinking demands that we should know which is which. Unless we can distinguish between the white filaments and the black the loom of thought will fail to weave a pattern which has either beauty or faithfulness to fact. To change the metaphor, we must pull down before we can build up; we must remove the rubbish before we can rear an edifice which stands out straight and square against the sky. As we must have light and shade in a picture, the darks showing up the lights and the lights showing up the darks, and each tone serving to define and limit another tone, so in thinking and speaking and writing we must deny as well as affirm. And our

negations are as useful as our affirmations. Thus every writer who aims at clearness has a leaning towards the antithetical style. Instead of painting his picture in grey and graduated tones he effects sharp contrasts. He does not limit himself to assertions, nor does he separate his denials from his assertions by a neutral zone of qualified statements. Placing his denials close up against his assertions, he gets clear edges to his ideas and point and piquancy into his style. The discerning reader need not seek far for examples: every branch of literature (and other printed matter) will supply them in abundance. Poetry, the most ancient form of literature, is full of negative propositions:

> Not all the waters in the rough rude sea
> Can wash the balm from an anointed king.
>
> Thou wast not born for death, immortal Bird!
> No hungry generations tread thee down.
>
> No useless coffin enclosed his breast,
> Not in sheet or in shroud we wound him;
> But he lay like a warrior taking his rest,
> With his martial cloak around him.

I open at random Emerson's essay on Compensation, and the first words my eye alights on are these:

> We cannot part with our friends. We cannot let our angels go. We do not see that they only go out that archangels may come in.

Emerson does not stand alone. Any other writer will supply abundant examples. The public speaker, too, knows the value of denials. By the judicious use of negatives he prepares his hearers' minds for the reception of positives. A certain distinguished educator thus began an address on *The Real Child*:

"Not the ideal child, not the perfect child, not the super-normal child, not the psychologist's child, not the Montessori child, not the Freudian child, but the real child—the child you and I meet in our homes and in the school." Many more of his sentences followed the formula: Not A nor B, but C. And, judging by the response of the audience, a very effective formula it was.

Turning to the greatest literature of all, we find such sentences as these:

> Charity envieth not; charity vaunteth not itself, is not puffed up, doth not behave itself unseemly, seeketh not her own, is not easily provoked, thinketh no evil; rejoiceth not in iniquity, but rejoiceth in the truth.
>
> For I am persuaded that neither death nor life, nor angels nor principalities nor powers, nor things present nor things to come, nor height nor depth, nor any other creature . . .

But why go on? In the face of these imperishable things are we to be scared by the maxim of a Victorian method-master?

The true-false test violates the maxim more flagrantly than these examples suggest. The maxim forbids us to exhibit an inaccuracy to the pupil, even if we at the same time tell him that it is an inaccuracy. The true-false test does not tell him: it gives no help in his task of sifting the true from the false. This is just what happens to us in ordinary life. We constantly come across true-false tests without knowing it. The daily paper is one of them.

CHAPTER IX

SPEED AND POWER

WHEN a man first meets an extreme example of the new examination, and realises the shortness of the time allowed, he receives a mild shock. He figures to himself a perspiring candidate tearing breathlessly through the paper, revealing nothing of his capacity to think, nothing of his capacity to weigh pros and cons, nothing of his capacity to organise ideas—revealing naught in fact but confusion and fluster. The whole thing seems a mad scramble for marks. The prototype of the new examination (the American Army Examination) is specially exposed to this sort of criticism; for only 23 minutes and 15 seconds were allowed for answering the whole of the 212 questions.

It is not surprising therefore that the question of a time limit should have given rise to abundant criticism. Those who object to a time limit generally do so on *à priori* grounds: their argument begins with the phrase "it stands to reason"—a kind of argument which it is always wise to regard with suspicion. One obvious fact is that every examination has a time limit. The candidates are not permitted to sit over this work indefinitely: the papers are rigorously called in at the expiration of a fixed time. It is only when the time limit is

narrow that doubt creeps in—when it is so narrow that the main determinant of success seems to be not the efficiency of working but the speed of working.

What the examiner wants to do is to differentiate the candidates—to spread them out—to bring the best to the top, leave the mediocre in the middle, and force the worst to the bottom. And there are two differentiating factors, time and difficulty. Just as in running a race. The competitors may run on a short level track, where winning is purely a question of speed. For in this case anybody could reach the winning-post in time—even the halt and the lame. But there is another, and a fundamentally different, mode of measuring one's ability to run. A track may be chosen which, instead of lying on level ground, would run up a mountain side and get more and more difficult with each foot of ascent—the ground rougher, the undergrowth thicker, the obstacles to progress more and more serious. Here a wide limit of time may safely be allowed. The difficulties of the course would suffice to spread the competitors. In both the level race and the mountain race we should be able to rank the candidates in order ; but while one would be an order of speed, the other would be an order of power—of endurance and skill.

As in testing the capacity to run, so in testing the capacity to think. If we construct a mental arithmetic test of 30 questions, all of approximately the same difficulty, we provide material for a speed test. The score is given by the number of items correctly worked in a brief fixed time, so brief that nobody, or at least not more than one, is able to

finish the paper in time. Now suppose we construct another mental arithmetic test beginning with easy examples and gradually increasing their difficulty until we made the thirtieth so hard that no more than one candidate can cope with it. The time factor here is unimportant. The difficulty of the examples suffices to distribute the candidates and yield a reasonable order of merit. In a speed test the candidate works against a constant resistance: in a power test he works against a steadily increasing resistance. In the speed test he could score full marks, if he had the time; in the power test he could score full marks if he had the brains.

Now it is quite conceivable, though by no means credible, that both types of test would yield substantially the same results—that the first in the one would come out first in the other, the second, second and so on. And that would be the only ground on which the speed test could validly be used in preference to the power test. Fortunately we are not called upon to decide this difficult question, since it has no practical significance. For the new examination falls into neither of the two categories—not even the Alpha Army Examination, which is generally singled out as an awful example of speed testing. Each of the eight tests which compose the Alpha Examination is made up of a number of items of gradually increasing difficulty. Each is an obstacle race with a time limit. Each is both a speed test and a power test. And the time limits, brief as they are, were so fixed that a small percentage of the candidates were able to finish the test in time. So the only pertinent

question is whether, if the same examination were set a second time to the same candidates and a much longer period allowed, there would be any serious differences in the two orders of merit. And this is a question on which we have trustworthy evidence.

The first important investigation into this matter was made during the war by Dr. Mark A. May under the direction of Professor Terman.[1] The Alpha Examination was given twice to the same recruits, to the number of 510; once with the regulation time limits, and once with double the regulation time limits. And the correlation between the two results was found to be ·965. This is extraordinarily high; higher in fact than the ordinary coefficient of reliability—higher, that is, than a test usually correlates with itself. This means that it does not seriously matter whether we allow five and twenty minutes or fifty minutes for the working of the army tests, provided, of course, the same conditions are imposed upon all. Whichever period we adopt, the candidates will be distributed in virtually the same way. Their marks will be higher, but their relative positions will be the same.

Another experiment has recently been made with the same test material, but on very different human material. It was carried out by G. M. Ruch and Wilhelmine Koerth upon 122 freshmen at the State University of Iowa.[2] All the freshmen

[1] *Psychological Examining in the United States Army, Memoirs of the National Academy of Sciences*, vol. xv, 1921, p. 416.

[2] *Journal of Educational Psychology*, vol. xiv, no. 4, April 1923, pp. 193–208.

had been classified on the basis of the combined scores of four distinct intelligence examinations; and for this investigation a selection was made of the best candidates (52), and the worst (70). The average candidates were left out. The following brief description of the three stages of the experiment may prove useful to the readers who wish to repeat the investigation with other test material. In the first stage the examinees were given Form 7 of the Alpha Examination under the strict rules laid down in the *Examiner's Manual*. Lead pencils were used, and were collected at the close. In the second stage blue pencils were distributed, the tests were repeated under the same time limits, and the candidates were allowed to make what additions or corrections they pleased, provided they did not obliterate any of the lead pencil marks. The blue pencils were then collected and a few minutes' rest allowed. In the third stage red pencils were given out, and the candidates continued working until they felt sure they could do no more. They then handed in their papers.

It was thus possible to find out precisely what each candidate could do in single time, in double time, and in unlimited time; to draw up three distinct orders of merit; and to judge how these three orders tallied with one another. Again the correspondence was extraordinarily close. The correlation between single time and double time was ·966, and between single time and unlimited time ·945. The correlations were high, not only for the whole examination, but for each of the eight tests taken separately. With one exception. The third or common-sense test, which required the

candidate to select the best of three reasons for a statement, yielded somewhat different results the second and third time from what it did the first time. The correlation between single and double time was ·455, and between single and unlimited time ·271. In the army experiment, however, it was as high as ·879. With this single exception the correlations found in the army experiment and in the Iowa experiment are in striking agreement. And they all point to the conclusion that the prominence of the speed factor in the army tests does not invalidate the tests. There is abundant evidence to show that speed and power tend to go together.

The question now arises: Who gains the more by an extension of the time, the dull candidate or the bright candidate? Since the correlations between the different results are so high, it is clear that one group does not gain any serious advantage over the other. An analysis of the figures shows that the better candidates gain the more absolutely, and the worse candidates relatively. In the army experiment the average score was 62, in the Iowa experiment 127·6. When the time was doubled the army average was raised to 80·5 and the Iowa average to 149·6. While the duller group gained 18·5 marks, the brighter group gained 22; although when considered as percentages on the original scores the gains work out as 30 and 17 respectively. It must be remembered, however, that the Iowa group were working at the top end of the scale, where there was comparatively little room for improvement. If they had had the same headroom as the army group they would have gained

much more. A comparison of the two groups, the high and the low, in the Iowa experiment leads to the same conclusion. A curtailment of the time imposes no special disadvantage upon any general type of candidate.

In spite of all this we must admit that a narrow time limit leaves in the mind of the candidate an uncomfortable feeling, a sense of dissatisfaction, a stubborn conviction that he has not done himself justice. He feels sure he could have done better if he had been given a little longer time. And all this is true: he would have done better. But since every other candidate feels the same, since all have been subjected to the same handicap, nobody should object to having the race arrested when it is in full swing; especially when it can be shown that the relative positions of the runners when they are so stopped is as good an index of their running powers as any other that has been devised. It can, at any rate, be shown that for a vast, an overwhelming majority, short-period testing, when properly carried out, is as sound and as valid as long-period testing. There may be—there probably are—isolated instances where this is not true. It would be a profitable piece of work to find these instances and to study them carefully. How numerous are they? And what is the special type of mind? Does it work fitfully, or with a plodding slowness? Are there emotional factors to be taken into account? At present we know none of these things.

We know, however, that the popular belief in slowness of thought being allied with profundity of thought must go the way of two other popular

beliefs which have already been discredited by psychological research—the belief that the slow reader gets more out of his reading than the rapid reader, and the belief that the slow learner retains what he learns longer than the rapid learner.

CHAPTER X

A WAR OF WORDS

A KEEN controversy has recently been raging in America over mental tests. It began in 1922 at a Teachers' Conference at Chicago, where Professor Bagley read a paper on Educational Determinism. Up to that time the mental tester had had it pretty much his own way. For so impressed was the American public by the success of the army tests that the utterances of the psychologists who devised the scheme were listened to with great respect. And the most astounding, not to say humiliating, statements were accepted almost without protest. The nation was induced to believe that the average mental age of the American citizen was only a little over thirteen years, that one-third of the adult population had a mental age below twelve, that the best brains of the army were to be found among the engineers, and that the younger the army officer the more intelligent he was.

One result of this confidence (or, if you like, credulity) was that an epidemic of mental testing broke out all over the States. In every branch of human activity—in business, in industry, in education—mental tests of all kinds were hurriedly invented and hurriedly applied. The testers were counted by the thousand; and, in the realm of education, by the million. There was some talk of

applying the tests in the realm of politics—not to the candidate but to the electorate. Dr. Cutten, the President of Colgate University, in his inaugural address, proclaimed his opinion that the suffrage should be based on mental tests. He declared democracy a failure and adult suffrage an absurdity; for, as he maintained, one-fourth of the free and independent citizens of the great republic hadn't the brains to understand the meaning of the vote. America could stand much from her professors, but she could not stand this; and the storm broke. But this was after Bagley had started a smaller storm within the professional camp.

What Bagley attacked was the determinism of the mental tester—his doctrine of the heredity of intelligence, and the consequent view that Nature settled all the important things in a man's life, and that Nurture had virtually no hand in the matter. If intelligence is the most vital factor in the mental make-up of a child, if intelligence more than anything else determines what he shall acquire at school, and what he shall achieve in life, and if this same intelligence is born with him and is unaffected by the early training in the home, by the later training in the class-room, and by the stern discipline of experience, then what is there left for education to do? If a man can never be an original utterance but is always, as Emerson puts it, "a quotation from his ancestors," then the school cannot touch the deeper and more permanent things of the mind; it can only deal with the superficial, the trivial, and the transitory. It cannot mould the essentials: it can merely mould the accidentals.

The attack on the hereditary view of intelligence

was taken up with much vigour by Mr. Walter Lippmann in *The New Republic*. Mr. Lippmann has no patience with the apostles of heredity, and he boldly tilts at one of their most impregnable examples—the record of the Kallikak family which was traced by Goddard in 1912. No modern book on heredity omits this story of the Kallikaks. Martin Kallikak, to use his accepted pseudonym, was a young soldier in the American Revolutionary Army, who had an illegitimate son by a feeble-minded girl whom he found in a tavern. After the war Martin married respectably, settled down, and reared a family. He thus became the progenitor of two lines of descendants, who number up to the present time about 500 on each side. And the contrast between the two lines forms an instructive study. The legitimate lineage is quite normal, with a fair sprinkling of doctors, lawyers, educators, judges, and prosperous men of business. But the other lineage, so far as it can be traced, seems to consist almost entirely of idiots and criminals. Confronted with these two genealogical trees, one manifestly sound, the other manifestly rotten, Mr. Lippmann is still unconvinced. While reluctantly admitting a possible taint in the pedigree, he maintains that conclusive evidence would require that the respectable line should have sprung from the tavern and the degenerate line from the decent home. As it is, it is not entirely clear to him whether the bad family history is due to a bad social start or to defective germ-plasm.

Mr. Lippmann is happier in his second argument. He points out that, according to the teachings of the intelligence tester, the rate of mental growth

declines as the mind matures. Rapid at first, it gradually slows down, and finally stops somewhere about the age of sixteen. It follows that the mental development that takes place during the first few years of a child's life is great in amount and profound in significance. Yet Professor Terman does not begin his observations till the child is four; and he has no right to generalise about the hereditary factor or data obtained after experience has been beating on a plastic mind for four years. The constancy of the intelligence quotient after this age may, Mr. Lippmann thinks, just as reasonably be ascribed to the persistence of the effect of the first four years' training as to some intrinsic quality inherited from the parents. At any rate, it is not fair to lump together the effect of natural endowment and the effect of infantile education and to ascribe the whole result to the germ-plasm.

And now Professor Dewey comes along, pats Mr. Lippmann on the back, and starts a new offensive in a series of articles entitled, Individuality, Mediocrity, Equality, and Superiority. One of Dr. Dewey's most noticeable traits is his dislike of uniformity, except in the matter of long words. It would evidently irritate him beyond measure to see a row of houses all of the same kind, or a crowd of people all dressed alike, or a class of children all doing the same thing at the same time. The dead level is to him anathema. He is all for oddity, uniqueness, individuality. On this score he vilifies the concept of mental age. To say that a boy born in 1912 has in 1922 a mental age of ten years only means that he belongs, on the basis of his performance of certain exercises, to a class of persons,

over a million strong, who were born in 1912. This, says Dr. Dewey, throws precious little light on the lad's intrinsic capacities. He does not object to statistical classification: it is necessary as a preliminary; but only as a preliminary. He does not blame the mental tester for starting with the mental age: he only blames him for stopping there. For the important thing is not to find out how a particular child resembles other children, but to find out how he differs from other children. In other words, Dr. Dewey pleads for analytic, diagnostic, or vocational tests.

The reply of the mental tester is that by far the most significant vocational test is the intelligence test. For the most serious misfits in life are not the round pegs in square holes nor the square pegs in round holes, but the big pegs in little holes and the little pegs in big holes. Either the man is too big for his boots, or his boots are too big for him.

Dr. Dewey girds at the psychologist's habit of labelling people "superior" or "inferior" on the basis of intelligence tests. For the term "superior" is meaningless in itself. It refers to some specific outcome. A person who is superior in making money is very different from the person who is superior in making furniture; and more different still from the person who is superior in making poetry. Dr. Dewey's argument wins from us an adventitious support from the mere fact that on this side of the Atlantic the word "superior" has acquired a sinister meaning, so that a very superior person is scarcely distinguishable from a beastly superior person.

This controversy, like many other controver-

sies, springs from a failure to discriminate. The disputants either cannot or will not distinguish between three separate meanings of the word "intelligence."[1] The word may, in the first place, mean that pure inborn ability to think efficiently which the psychologist is seeking, but has not yet found—the mother-wit that is unaffected by special circumstances and is neither sharpened by training nor dulled by schooling. Secondly, it may bear the popular meaning which covers all forms of intellectual adroitness to whatever cause it may be due. Thirdly, it may mean that group of intellectual factors which is measured by mental tests. And since each scale of tests measures its own group, this last meaning is as manifold as the scales themselves. The third kind of intelligence consists of a mixture of the other two in unknown proportions.

The critics find a base for their attacks in the pretence that the third meaning is identical with the first. Yet no psychologist of repute ever claimed that any of the tests in existence, either the Binet Tests, or the Army Tests, or any other tests, afford a mathematically exact index of pure native intelligence. The most they afford is presumptive and approximate evidence. The evidence is presumptive only because we cannot be sure that the testers have been in the past subjected to the same influences of environment and education. If they had been so subjected, any differences of achievement would be due to inherited qualities. But

[1] Those who are interested in the theoretical implications of the term "intelligence" are recommended to read Professor Spearman's recent book: *The Nature of Intelligence and the Principles of Cognition.*

since general intelligence is not the only inherited factor, we cannot even then be sure that pure general intelligence is being measured.

The critics do not of course doubt the facts of heredity. They merely doubt whether intelligence tests yield conclusive proof of the heredity of the mental traits which they probe and gauge. What the tests really reveal is a certain intellectual background; they fail to say with certainty how that background has come into being; they fail to separate the part that is inherited from the part that is acquired. They fail because the child who comes from a good stock also comes in ninety-nine cases out of a hundred from a good home. He not only derives from his ancestors; he lives with his ancestors—his more immediate ancestors, at any rate. The very people who endow him with his mental powers are also engaged, consciously or unconsciously, in cultivating those powers. Through themselves, their relatives, and their circle of friends they provide him with a social environment which is of a piece with his heredity.

Thus, on the vexed question of heredity, mental tests do not tell us much more than we knew before. Even the constancy of the intelligence quotient (if it really is constant) does not prove that intelligence is independent of training. For if, as one of the critics has suggested, children's weights were treated in precisely the same way as their intelligence, and weight quotients obtained by the same statistical methods, it would be fairly certain that a child's weight quotient would be about as constant as his intelligence quotient. Yet it would be folly to maintain that weight is unaffected by feeding.

So persistent is the confusion between the two meanings of the word "intelligence" which I have pointed out above—the professional meaning and the popular meaning—that it fogs the issue in every discussion of the problem that takes place. It is important to remember that the intelligence of which the psychologist speaks—the hypothetical mental factor that is unaffected by training, that grows with the child's physical growth, and matures in the early teens—is highly abstract and of meagre content. By saying that it is abstract I do not mean that it is unreal. The colour red is abstract: it has no independent being—no existence apart from the complex qualities which go to make up the red object. But red is real enough for all that. So with intelligence. It, too, is abstract in the sense that it does not stand out distinct and isolated, but is indissolubly bound up with a host of other mental factors. We can think of it by itself: we cannot find it by itself. And we can best get at its nature by gradually removing the other qualities with which it is liable to be confused. Let us do this. Let us strip it of its trappings and see what remains; or, which is the same thing, let us first find out what it is not.

It is not knowledge. A man may assimilate an encyclopædia, or pack his mind with the lore of many lands, without adding one jot to his intelligence. It is not habit, and has nothing to do with the ease that comes with practice. A man may acquire extraordinary facility in dealing with numbers, with words, or with certain groups of facts, without making his intelligence any brighter or any duller. It is not interest. A man may become

keen in the pursuit of knowledge—" may follow knowledge like a sinking star beyond the utmost bound of human thought "—or may develop an intense ardour for some branch of artistic activity, and leave his intelligence just as it was before. It is not the capacity for application—the capacity for taking infinite pains. That is a matter of character; and character has nothing to do with intelligence. All these adventitious things having been removed, the residue is the psychologist's " intelligence."

One source of difficulty is the fact that this abstract quality not only carries with it all these other qualities that we have just negated, but, as a general rule, carries them in a measure that corresponds with its own intensity. As a general rule, though not as an invariable rule, the more intelligent a person is the more familiar is he with the ways of men and of nature, the more interested is he in things of the mind, the more persistent is he in the pursuit of knowledge. He knows more, he can do more things, he is more at home in the world. And all these variable accompaniments are included in the plain man's meaning of intelligence and are excluded from the psychologist's meaning of intelligence. And this exclusion leaves nothing but a sort of general mental energy which is available for traversing old fields of thought and for breaking new intellectual ground.

The distinction between the two kinds of intelligence is somewhat analogous to the distinction between temperature and heat. A pint of water at boiling-point has the same temperature as a quart of water at boiling-point, but it has only half as much heat. Intelligence (in the technical

sense) is temperature and not heat: an intensive magnitude rather than an extensive magnitude. Binet had this distinction in mind when he spoke, as he constantly did, of different levels of intelligence instead of different amounts of intelligence.

Bagley makes a similar distinction between the "vertical growth" of intelligence and the "horizontal growth" of intelligence. And he attaches greater importance to the horizontal. He writes thus: "I say that Darwin and Lincoln could do a type of thinking at fifty that they could not do at twenty-one—or at thirty-one, for that matter. I contend that the growth in the intervening years was real intellectual growth, and if it be objected that it is not growth in 'general intelligence' I answer that it then represented something vastly more important than general intelligence."[1] This is clear. Whether we agree with him or not we at least know what he is talking about. I will merely remark that if Professor Bagley were suddenly translated to the planet Mars (I should be sorry for it to happen) he would find his vertical intelligence of far greater use to him than his horizontal.

Many a man has an uncomfortable feeling that mental tests will end in sticking labels on people. He has visions of men going about with their breasts decorated with intelligence quotients; or carrying in their pockets certificates to show that they have been examined by registered psychologists and are certified to have highly superior brains. And the fool is to be given the bad name which goes a long way towards making him deserve it. These are

[1] *Journal of Educational Research*, vol. iv, no. 5, December 1922, p. 381.

dismal and desolating prospects it is true; but there is not the least likelihood of their being realised. The intelligence quotient is recognised as merely a rough and tentative gauge of a pupil's educability: it is a help to the teacher in finding the scholar, not to the scholar in finding his place. And the teacher keeps the intelligence quotient to himself: it is as much a secret as the doctor's diagnosis which enables him to heal his patient. Even if the pupil is told his intelligence quotient (it is contrary to professional etiquette) he will not understand what it means; and if he understands what it means it will disturb him no more and no less than his examination marks. In fact, mental testing has no more power to brand or hall-mark than has the ordinary examination. What little power it has is invested in it by the law of the land and is put to beneficent use. The law demands that children who are to be educated as mentally defectives should be certified by a qualified medical man, and the certification has been made with incomparably greater justice since Binet invented his scale. To the afflicted, mental testing is more of a protection than a menace.

To those at the other end of the intellectual scale—to the highly intelligent—the university degree is a sufficient label. With it the intelligence quotient is in no way likely to compete.

Some of the recent researches of Mr. Hugh Gordon[1] have shaken our belief in the validity of the Binet tests as a scale for all and sundry. He has been applying the Stanford Revision to physically

[1] See *Mental and Scholastic Tests among Retarded Children* Board of Education.

defective children, who attend school infrequently, to gipsy children, who attend less frequently still, and to barge children, who attend least frequently of all. And he has found that the intelligence quotients of these children tend to rise and fall with their attendance at school; and that the intelligence quotients of gipsy children and barge children get lower and lower as the children get older. These facts are highly significant. They either show that Binet's scale is valid for school children only, and do not gauge the intelligence of children whose minds are not exposed to the influence of teachers and books; or else they show that intelligence if neglected tends to wither and fade. Probably both conclusions are justified: each fact contributes something to the total result. Indeed the careful researches of Dr. Burt should have led us to expect a failure of the Binet tests when applied to out-of-school children.

Another significant fact revealed by Mr. Gordon's inquiry is that scholastic tests, if devised and applied with the same care and precision as intelligence tests yield very similar results. This is only true, however, when the testee has attended school long enough to have begun to read and cipher. The scholastic tests used by Mr. Gordon were three of my simple oral tests—the one-minute reading test, the one-minute addition test, and the one-minute subtraction test.[1] He found that with the special children in question the educational age as determined by the three tests was practically identical with the mental age as determined by the Binet tests.

[1] See pp. 145–147.

It would be idle, therefore, to maintain that mental tests as they exist to-day finally settle any of the theoretical issues of psychology or education. They do not succeed in dissecting the mind so as to separate innate abilities from acquired abilities, or to separate general ability from specific abilities or to separate mother-wit from book-learning. Ideal mental tests would do these things; actual mental tests, if they do them at all, do them very roughly. This every mental tester is ready to acknowledge; and he does not care a dump if the whole fabric of biological and psychological theory that has been prematurely built up on mental tests comes crashing to the ground. If it does, may it fall upon the doctrinaires who built it up. "And thereupon the Pedant shall sit upon the Abstract Bagman, crushing him; and the Pedant shall choke in his own fat." The mental tester does not care because, whatever happens to the superimposed theory, mental tests remain as they are—an inestimable boon to the teacher who wishes to discover the educability of his scholars. If the teacher can map out the mental background of each of his pupils he does not worry about the source of that background. It does not matter to him whether it is due to parentage or to tutorage so long as he can feel sure it is there as a stable thing upon which he can rely. The ordinary examinations do not give him this assurance; mental tests do. In fine, the aim of the mental test movement is not so much to prove theories as to improve examinations.

CHAPTER XI

FLECKS AND FLAWS

THAT the new examination has apparent flaws its most ardent advocate will admit. And all he can do in defence of his position is to show that the flaws are not real; or, failing this, to show that they are not so great, nor yet so serious, as they seem to be.

The most palpable defect (the defect that the critics most eagerly pounce upon) is the absence of any means of testing the mind's capacity to cope with big things—to organise ideas, to construct large and complex wholes. The new examiner is accused of dealing in scraps and snippets, and of pretending that shreds and patches, if they are numerous enough, will make a complete garment. The critics make much use of the term "organic whole," and they reproach the new examiner with ignoring its significance and its claims. They point to the essay, the long mathematical problem, and the full answer to a question in geography, history, or science; all of which they laud as fine examples of the organic whole. And they maintain that without such examples of the organic whole no examination is worthy of the name.

The new examiner's reply is twofold. He first asserts that an organic whole is of little use to the examiner if he cannot measure it; he then main-

tains that the thing he himself ultimately measures —not ostensibly measures, but ultimately measures —is just as much an organic whole as the thing measured by the old examiner. The new examiner marks a single word or a single figure: the old examiner marks a multitude of words and a multitude of figures. It is true that these two things are poles apart. But these are not the real things that the examiner is trying to evaluate. They are but outward tokens, but signs and symbols of the inner thoughts—of the ability to think to good purpose. And it is upon this ability to think that the examiner in the last resort pronounces his verdict.

We touch here the heart of the matter. For the crucial question is: Are the thought processes with which the new examiner deals different in kind as well as in bulk from the thought processes with which the old examiner deals? Are they a different sort of thing, or are they merely more of the same thing? The reply of the new examiner is that they are merely more of the same thing. He deals with a smaller unit but not with a unit of a different quality. To supply a missing word in a passage of prose, to straighten out a mixed sentence, to extend a number series, to select the right reason from a number of suggested reasons—to do any of the apparently simple tasks imposed by the new examiner brings into play the same complex trains of thought as are evoked by the broader questions of the old examination. The examined products are different, but the evaluated processes are the same.

It is not easy to believe that the evaluated processes are the same: it is not easy to believe that

the inner organisation of the smaller unit is as complex as the inner organisation of the larger unit. For the larger unit includes the smaller units and includes them in a fixed and definite order. There is a superimposed integration of elements. It has, however, been claimed that there is in the smaller unit an integration of still smaller elements—a noetic synthesis, as Stout calls it—which, though less consciously exercised, is an index of the same capacity as that which underlies the larger synthesis. We here reach a deeper level and see that when the new examiner claims an essential similarity between the two processes he places the real identity in the mental capacity that underlies both. The proof of this identity lies in the realm of statistics. The proof is not as complete as it might be; but there is already enough evidence to render it a tenable theory that the capacity for organising knowledge is as adequately gauged by the new examination as by the old.

It is a tenable theory, but not a demonstrable theory; and it is possible that in maintaining it the new examiner is pressing his point too far. It will need very strong evidence to convince anyone that it requires no higher ability to write an essay than to cope with a number of simple, direct questions which collectively cover the same ground. and separately are answerable by a single word. Till this evidence is forthcoming—if it ever will be forthcoming—the new examiner is well advised to test directly the capacity to arrange and combine the smaller units so as to form the larger units. He already does this, as a matter of fact, in some of his tests. Picture tests in which a number of

drawings representing a series of events have to be placed in proper order are by no means uncommon. The sequence demanded is generally a time sequence; but there is often a logical relation as well as a temporal relation between the several elements. This is specially true of my test in Constructive English, which appears on pp. 176–178. More tests of this type may readily be constructed. As a simple instance, we may require the following items, which represent an example in the addition of fractions, to be arranged in their proper order:

$$5 + \tfrac{3}{6} + \tfrac{2}{6} = 5 + \tfrac{1}{2} + \tfrac{1}{3} = 5\tfrac{5}{6} = 3\tfrac{1}{2} + 2\tfrac{1}{3}$$
$$= 5 + \tfrac{5}{6} = 3 + 2 + \tfrac{1}{2} + \tfrac{1}{3}$$

It is not therefore strictly true to say that the new examiner fails to test the power to organise knowledge. What he really fails to test is the power to initiate trains of thought and to express himself in words. And it is idle to pretend that this is not a serious defect. The new examiner cannot measure originality; and he cannot measure felicity of expression. And he does not pretend to. But he is not very unhappy about it. When the old examiner reminds him that he cannot do these things he turns to his critic and asks: Can you?

The new examination is, by a strange irony, open to attack from another quarter on the very ground of its efficiency. While examinations allowed of a wide margin of uncertainty—while they showed manifest signs of human fallibility—they were not taken too seriously. "Ah well!" says the father, "what does it matter if the lad cannot pass examinations? I was just the same myself." And in the tone of his voice there is a large implication of

future greatness for his boy. Examinations have in the past been treated lightly by practical men. But now that they bid fair to become efficient, now that they seem able to do what they purport to do, people are beginning to be alarmed. In the intelligence test they see special danger. One would imagine that the first to take the alarm would be the Tory—the man who, according to Carlyle, believes in the doctrine of standing still; for he would foresee the possibility of an aristocracy of talent supplanting the aristocracy of wealth and position. As a matter of fact, however, it is King Demos who is the first to be roused. People have begun to see in mental tests a menace to democracy. It is this fear that is really at the root of the controversy described in the last chapter. The real question at issue is sociological rather than biological. It is not so much whether the sense that John Smith now has is the sense he was born with or the sense he acquired, as whether John Smith has the right, as Bagley puts it, "to share in the spiritual life of the race." There is a lurking fear that mental tests may be used for robbing a large section of the community of a liberal education on the grounds that it is incapable of benefiting by a liberal education; that they may be used for effecting drastic economies in education on the plea that we are now educating the uneducable, or are educating children beyond their intellects, or are giving them an expensive education when a much cheaper education would do just as well.

There is some ground for this fear. But it is not the first time that what was intended as a tool has been used as a weapon—that what was designed for

beneficent purposes has been employed for harmful purposes. Mental tests were devised as means of interrogating nature—human nature—means of arriving at certain facts concerning the individual mind. They reveal in that mind certain limitations, but they also reveal certain potentialities, certain possibilities of culture. And the second revelation is not only more important than the first, but is surer and more trustworthy. Mental tests speak in clearer and more confident tones when they tell us what a child can do than when they tell us what he cannot do. Those, therefore, who seize upon the limitations and ignore the possibilities make an injudicious, not to say an illegitimate, use of the tests.

What precisely can mental tests do? They can enable us to arrange roughly a group of children in order of general intelligence, the term "general intelligence" being interpreted in a technical sense. They cannot tell us where falls the line of demarcation between the normal and the subnormal, or between the normal and the supernormal. They cannot tell us which child should be educated up to six years of age, which up to sixteen, or which up to sixty. They can tell us little, in fact, about the limits of a child's educability, except perhaps in the case of manifest idiots. If they point to a boundary in one direction they point to endless expanses in other directions. It is believed by many that the mental tester can place a boy in the category of mentally defectives with the same certitude as that with which the ordinary man can identify a a sheep or a camel. In point of fact, mental deficiency, as the term is at present used in education,

is, as Dr. Burt has shown us, a matter of school accommodation. In London, at the present time there is sufficient room in the special schools to accommodate about 1½ per cent. of the total number of school children. This 1½ per cent. is labelled mentally defective. If there had been accommodation for 5 per cent., 5 per cent. would be mentally defective. It was not mental tests that fixed the 1½ per cent.: that was fixed before mental tests were invented.

Although mental deficiency is at present a matter of accommodation, it should not really be a matter of accommodation. It should be a question of the pupil's ability to profit by the instruction given in the ordinary school. If the type of instruction is changed, so is the margin of mental deficiency. From the standpoint of the university as providing a normal education for people over eighteen years of age, the proportion of mentally defectives is appalling to contemplate.

It will thus be seen that what mental tests can tell us is always relative to the type of education given in our schools. If the bulk of the child population could not profit by the education given in our secondary schools, then the proper course would be, not to exclude the children, but to alter the curriculum. And in the task of adjusting the curriculum to the child mental tests can be of signal service.

Rightly used and rightly regarded mental tests serve large and noble purposes. Far from being hostile to, they are distinctly friendly towards, democratic ideals. For if they tell us anything at all they tell us that intelligence is so distributed

among a nation that the majority may be considered normal specimens of mankind, who are capable of entering into the greater part of the social and spiritual inheritance of the race. They tell us, in fact, that a liberal education is not a thing for the few but for the many.

CHAPTER XII

SETTING AND MARKING

THE old examination is easy to set and hard to mark; the new is hard to set and easy to mark. By setting an examination I do not mean the mere giving of the tests, but rather the inventing and arranging of them. In giving an examination which is already standardised one has only to obey printed instructions, and in marking the answers one has only to compare them with a printed key. The difficulty of setting an examination confronts the teacher when he has to construct his own tests; and this necessarily happens in the examination, casual or routine, of the class-room—the *ad hoc* examination—the examination that is specially designed to fit a specific syllabus and to gauge proficiency in a specific course of study.

The usual plan is to set an examination either at the end of the course of study, or at the end of a section of the course. And the teacher begins to think of the questions after he has finished with the teaching. This does well enough for the old examination, where the questions are few and vague; but it will not do at all for the new examination, where the questions have to be numerous and definite, have to cover the whole field of study, and have to touch upon everything that is im-

portant and upon nothing that is trivial. The teacher who wishes to examine by the new methods will, if he is wise, prepare his examination when he is preparing his lesson or his lecture. And he will prepare it as carefully as he compiles his notes. Indeed, his notes should form the backbone of his examination; for presumably they sum up all that is important in the facts and principles which he wishes to inculcate upon his pupils.

Let me illustrate by a concrete example. Let us suppose that a lecturer in Psychology is preparing a lecture on the functions of the brain, based on such material as is found in the earlier chapters of William James's *Principles of Psychology*. He is almost certain to include in his notes Meynert's terse dictum: The cortex of the brain is the surface of projection of every muscle and every sensitive point of the body. Here is matter for abundant exposition and illustration. Here, too, is the framework of a test which will reveal whether the theory has been grasped. The test may be put in the completion form thus:

The . . of the brain is the surface of . . of every . . . and every . . point of the body.

Or the test may be cast in the true-false form, thus:

(1) Every muscle in the body is anatomically connected with the surface of the brain.

(2) The cerebral cortex consists of distinct organs, each of which subserves a distinct faculty, such as memory, self-esteem, reverence.

(3) The brain is solely the neural basis of thought, and the spinal cord solely the neural basis of movement.

(4) The cortex is made up of certain nervous arrangements which represent movements and impressions.

(5) The higher nerve centres are found deeply embedded in the brain.

(6) The brain is an organ for controlling the movements of the body.

Or, again, the test may be set in the limited option form. Thus:

Of the following statements select the true one:

The cerebral cortex is
- (*a*) Concerned exclusively with thought processes.
- (*b*) Connected anatomically with the sense organs only.
- (*c*) Connected anatomically with the muscles only.
- (*d*) Connected anatomically with both sense organs and muscles.

No doubt many other, and better, tests will occur to the expert on the subject. I submit mine merely to show how the lecturer may develop his tests out of his notes. There is reason to believe that the very thinking out of the tests before the lecture will help to make the lecture clear, interesting, and incisive.

In constructing a true-false test special care is needed. I have found it a good plan to make all the statements true to begin with, and then to alter those that readily lend themselves to a false presentation. One should end up with about the same number of each. This is to prevent the candidate from detecting a bias in the teacher's questions towards either the true or the false form, and from guessing on the line of the suspected bias. On the other hand, the number should not be exactly alike, else the shrewd candidate will count his answers and try to make the true and false

balance. Sometimes there should be more true statements than false; sometimes more false than true. They should not appear in strict alternation; nor, in fact, in any uniform order. The only regular thing about them should be their irregularity. They should have the element of surprise that belongs to chance.

I have sometimes found my false statements to be useless as tests through being too obviously false. They were so flagrantly false that they deceived nobody, and left the final distribution of marks unchanged. The best kind of false test is one which is liable to entrap the candidate who does not know his subject, but cannot possibly entrap the candidate who does. Take, for instance, statement (5) above: "The higher nerve centres are found deeply embedded in the brain." This would seem quite plausible to the man who knew no brain physiology, and it would readily gain his assent. And he need not be regarded as a fool. But he would be almost impossibly stupid if he swallowed this statement: The higher nerve centres are to be found in the left elbow.

Another good form of false test attacks discredited, but not discarded, theories. The second of my sample statements ("The cerebral cortex consists of distinct organs, etc.") embodies a belief still held by phrenologists.

A statement intended to be marked as false should be simple and direct. It should, in fact, be one statement and one only. If I say: "Water freezes at 32° F., and boils at 100° F.," I make two statements, one of which is true, and the other false. To mark the sentence definitely one or the

other is impossible. So with this sentence: "Ice floats on water because it is a solid." Here we have a statement which is true, and a reason which is false.

Sentences of this latter type are sometimes useful to test reasoning in geography, history, or science. If they are included, a note should be inserted at the head of the test to the effect that no item is to be regarded as true unless it is wholly true.

Before starting a new examination the answer papers should be got ready by the pupils, who should put down their names, ages, and standards, and insert the numbers of the questions. The arrangement of the paper should accord with some such plan as is suggested by the various keys that appear in this book. One of the advantages of this scheme is that none of the actual time of the examination is spent on accessories: all the time is devoted to winning marks. Another advantage is that it enables the papers to be marked by stencil.

To the question: who should mark the papers? the answer is, the pupils. If the examination is competitive and the future careers of the candidates are determined by their success, then, of course, an independent marking is essential. But for all the ordinary purposes for which examinations are held in schools, it is far better that the pupils should mark their own papers. It is certainly more trustworthy. Some time ago I tested the intelligence of a class of boys by means of a group test. The test was finished just about the time for dismissal, and I suggested to the teacher that I should return the next day and get the boys to mark the papers. He protested. The boys couldn't be trusted to do

this. He believed in doing a thing of that kind thoroughly. So he always did it himself. If I did not mind, therefore, he would take the papers home and do them himself. I did not mind. He also rejected the idea of a stencil. He was a man of experience and had his own way of doing things. He accordingly took the papers home with him and brought them back the next day all marked. But when I checked the marking I found scarcely a single paper marked correctly. At another class in the same school where the boys marked the papers I failed to find more than one blunder in the whole set; and that was a perfectly plausible one. An answer had been written illegibly and had been misread.

In William de Morgan's novel, *Joseph Vance*, the hero's father had a favourite motto to the observance of which he ascribed his success as a builder. Here it is: If you want anything done well, don't do it yourself. Dangerous as this maxim would be in the hands of the injudicious, it will prove quite a sound piece of advice for the scoring of the new examination. Don't mark the papers yourself; let the pupils mark them. And let the marking follow close upon the heels of the examination. If any interval at all intervene let it be the briefest possible. For the main advantage of self-marking is that the test is thereby converted into a lesson. The pupil ends up knowing more than he did when he began.

There are various ways in which the pupils may mark their own papers. While the teacher calls out the answers each pupil may mark his own, or each may mark somebody else's. Perhaps the best

plan of all is to call the answers out twice, once for the marking by each pupil of somebody else's paper, and once for the verification of his own marks. Any mis-marking is certain to be challenged and rectified. All this should be done while the boys' brains are warm on the work.

Children should be taught to make records and graphs of their own scores. They soon show an absorbing interest in their own progress and try to beat their own records. This self-competition has all the merits of competition, and none of its evils.

If the teacher decides to mark the papers himself, he would be well advised to use a stencil. The simplest form of stencil is the edge of a piece of paper. Let us suppose, for instance, that an arithmetic examination has been worked on foolscap and that the children have been required to arrange the answers in four columns, with 25 answers in each column. The only stencil necessary will be a sheet of foolscap with the first 25 answers arranged in a column near an edge of the paper, the next 25 near another edge, and so on. If both sides of the paper are used, and it is turned upside down for writing two out of the four columns, a stencil is provided by which edges may in turn be placed close up to the columns on the pupil's papers, and the answers checked very rapidly.

CHAPTER XIII

A TYPICAL NEW EXAMINATION

AMERICA has produced standard educational tests in such bewildering profusion that the plain Englishman is sorely tempted to have nothing to do with any of them. Rival publishers advertise them as though they were fountain-pens or safety razors. And, if we are to believe the picture drawn by Professor McCall,[1] to examine children by the latest methods is a terrific business. A host of examiners (some of them mere university students) pounce upon the school, each armed with his own packet of standard tests, and each prepared to use a highly specialised technique. In one subject only—that of reading—each child has to be given fourteen distinct and separate marks. The examiner must record the testee's chronological age, initial reading score, initial reading age, initial reading quotient, initial mental age, intelligence quotient, initial accomplishment quotient, estimated final reading age, estimated final mental age, reading score objective, final reading score, final reading age, final accomplishment quotient, and final accomplishment quotient minus initial accomplishment quotient.

This reduces the whole thing to an absurdity. Such a scheme is likely to be crushed to death under

[1] *How to Measure in Education*, pp. 67 ff.

the weight of its own machinery. It cries aloud for simplification—for a single unified scheme. Instead of using Mr. A.'s scale for Reading, Mr. B.'s for Arithmetic, Mr. C.'s for Spelling, and Mr. D.'s for something else, it would be better to devise one scale for all subjects. And this is what has recently been done by Professor Terman and his colleagues at Stanford University. Through the courtesy of one of the authors (Dr. Kelley) I have received a copy of the new *Stanford Achievement Test*,[1] by Truman L. Kelly, Giles M. Ruch, and Lewis M. Terman, which seems to be the simplest and most complete examination of its kind that has yet appeared. And it is an excellent example of the new examination. Although there are nominally two examinations, one primary and the other advanced, these examinations are really one, for the primary consists of a selection of the easiest questions in the advanced. So the advanced examination alone need be described.

It consists of nine tests, three of which deal with reading, two with arithmetic, and, as might be anticipated, none with composition. One of the characteristics of the new examination is the emphasis placed on silent reading. It rejects reading aloud as an examinable product as uncompromisingly as it rejects the essay. But whatever may be said about its rejections, its inclusions are commendable enough. If from among the many specific aims of schooling we had to select the most essential of all we should probably agree in singling out the right use of books. And in testing the capacity to use

[1] Copies may be obtained of the World Book Company, Yonkers-on-Hudson, New York.

books it is not enough to set the pupil to "bark at print" and then to record one's opinion of the bark; one must discover what meaning he can absorb, and at what rate he can absorb it. The essential thing is understanding, not delivery.

Of the threefold reading test the most interesting and the most original is the first, which is designed to measure the child's ability to grasp the meaning of large units. It consists of twenty-nine paragraphs each of which has one or more missing words. These missing words cannot be supplied by merely glancing at the immediate context: the whole paragraph must be read and understood. And if it is read and understood the missing word is inevitably found.

The first paragraph reads thus:

Fanny has a little red hen. Every day the hen goes to her nest and lays an egg for Fanny to eat. Then she makes a funny noise to tell Fanny to come and get the . . .

The last paragraph reads as follows:

"Naïve" and "unsophisticated" are frequently confused. The former suggests a type of behaviour which is artless, spontaneous, and free from the restraints of custom. The latter implies fully as great lack of knowledge of social usage, and, in addition, conduct which is primitive and perchance inelegant. Thus, the . . . youth was the first to enter the car, and his . . . little sister warmly kissed him in the presence of the king. We may also say that a country boy is . . . with respect to city life and customs.

This is a distinct improvement on the method of testing comprehension by getting the pupil to reproduce in writing the substance of what he has read; for the reproduction method tests memory as well as understanding. A child may clearly under-

stand the whole of a passage without being able to reproduce more than a fraction of it.

The second test deals with the meaning of sentences, and is of the true-false type. The first and last questions are :

1. Is milk white ? . . . **Yes.** No.
80. Might a hysterical person given to rashness be intolerable ? . . . **Yes.** No.

The third reading test has to do with the meaning of words. The pupil is required to underline the word that makes each sentence true. I quote the first and last items.

1. March is the name of a **day, food, month, week, year.**
85. Candid means **illegitimate, impeccable, imperious, incisive, ingenuous.**

The two arithmetic tests need not be described as they consist of ordinary brief examples in mechanical arithmetic and problems.

Test VI (Nature Study and Science) begins and ends thus :

1. Thanksgiving comes in **July, January, November.**
95. Fondant is a kind of **candy, meat, salad.**

The corresponding items in Test VII (History and Literature) are :

1. An elf is a kind of **animal, brownie, dragon.**
95. The singular of " are " is **is, was, were.**

In Test VIII (Language Usage) the pupil has to underline the correct alternative. I quote the first item and the last.

1. I **calculate / expect** to go soon.
60. Is that **he ? / him ?**

The last test of all is a spelling test.

The merits of the Stanford Achievement Test are many. It enables all the children from seven years of age to fourteen to be examined from the same booklet; it covers all the essential subjects of school instruction; it has been carefully standardised; and it gives the norms for each age-group and thus avoids the fault of other American achievement tests which give norms for the grades only. In spite, however, of its merits it cannot be applied to English children without drastic changes; for many of the items are strongly tinged with local colour. In the arithmetic tests there is no reference to pounds, shillings, or pence, but only to dollars, dimes, and cents. And some of the other tests presuppose a social environment which is alien to English children. If an English boy found in his science paper a question about thanksgiving he would be surprised. He would also be floored.

Useful as a standard test is in telling a teacher how his children stand in relation to other children, it is no less useful in suggesting new modes of examining and new types of questions. It can serve not only as a standard to which to conform (or from which to depart), but also as a model on which to mould the casual or formal examination of the class-room. It helps to improve the quality of home-spun and home-brewed.

CHAPTER XIV

WHAT TO DO

THIS is a chapter of practical advice. Many a head teacher feels moved to try the new testing but does not know how or where to begin. The best place to begin is with the children at the bottom of the school. Catch them as young as possible. The secret of good organisation is to put each pupil in the class where he will progress at his natural pace. And his natural pace is determined by his intelligence. It is important, therefore, to gauge the intelligence of the new scholars. The next best thing to an intelligence test is a carefully devised test of achievement. The worst of all is an old-fashioned examination. Perhaps the safest guide is a two-fold new examination, one part to measure the entrant's intelligence and the other to measure his attainments.

In the secondary school there is no difficulty; for there are available for winnowing the new-comers not only such group tests of intelligence as the Northumberland Tests, the Simplex Tests, the Chelsea Tests, the Columbian Tests, and the Civil Service Tests,[1] but also such attainment tests as appear in this volume and in Dr. Cyril Burt's book *Mental and Scholastic Tests.*

In the elementary school the problem is not so easy. The entrants to the infant school are so young that they must be tested orally and indi-

[1] See Ballard's *Group Tests of Intelligence.*

vidually and at some length. It requires patience and tact and no small measure of technical knowledge to discover what sort of intellects these immature children possess. But the discovery is well worth making. The children in infant schools are generally graded according to age: they should be graded according to intelligence. For wherever the grading has been based on intelligence it has proved to be signally successful.

As the most satisfactory test of the intelligence of very young children is undoubtedly Binet's, it is well that English teachers should know what translations and revisions of the Binet-Simon scale are readily obtainable in England. First and foremost there is the comprehensive revision made by Dr. Cyril Burt and published in his *Mental and Scholastic Tests* (King and Son). Then there is a less detailed form of Dr. Burt's translation which appears in my *Mental Tests* (Hodder and Stoughton). A still more concise version by Miss Margaret Drummond will be found in an appendix to *Mentally Defective Children* by Binet and Simon, translated by Dr. W. B. Drummond (Edward Arnold). A very good revision of the scale up to eight years of age made by Mr. W. H. Winch (Ralph, Holland & Co.) has the additional merits of being very cheap and very convenient. The well-known and admirable Stanford Revision appears in Terman's *The Measurement of Intelligence* (Harrap). Purchasers of this work must be careful to get the Test Material as well, as the book is incomplete as a practical guide without it.

To administer the Binet tests with a reasonable degree of precision requires some measure of

psychological training; and realising this, many teachers, perhaps with an excess of diffidence, hesitate to use them. I therefore suggest an alternative plan, which any and every teacher can put into practice. In my book *Mental Tests* will be found three very simple one-minute tests, one for reading, another for addition, and another for subtraction. I have used these tests extensively for the last ten years; and the more I use them the greater confidence I have in their value and their validity—a confidence which has been strengthened by the researches of Mr. Hugh Gordon, to which I have already referred. He found that for the lower mental ages these three tests combined yielded an attainment quotient which closely approximated to the intelligence quotient given by the Binet scale. There are, in fact, good grounds for regarding these tests as thoroughly sound so far as they go, and of great practical use in classifying and promoting all young children who have begun, in however rudimentary a way, to read and to cipher.

The reading test consists of a leaflet of simple words—a duplicate, with the numbers omitted, of the set of words given in the accompanying scales. The test copy is handed to the child, who is told to read the words as fast as he can. At the end of one minute the examiner stops him and records his reading age as given in the two scales, one of which is for boys and the other for girls. If the child hesitates for more than five seconds over a word he is prompted and told to pass on, that particular word counting as an error. The score is the total number of words correctly read in a minute—that is, the total number read minus the number misread.

One-minute Reading Scale (Girls)

		5-6	5-6	5-7	5-7	5-8	5-8	5-9	5-9	5-10	
is	me	on	at	by	so	us	an	it	or	be	11
5-10	5-11	5-11	6	6-1	6-1	6-2	6-2	6-3	6-3	6-4	
to	as	he	of	in	go	up	am	if	no	we	22
6-4	6-5	6-5	6-6	6-6	6-7	6-7	6-8				
my	ox	do	the	and	for	but	him				30
6-8	6-9	6-9	6-10	6-10	6-11	6-11	7				
are	can	she	dog	let	you	not	was				38
7-1	7-1	7-2	7-2	7-3	7-3	7-4	7-5				
out	try	see	mix	cat	now	boy	saw				46
7-5	7-6	7-7	7-7	7-8	7-8	7-9	7-10				
bit	met	top	run	man	pet	lot	get				54
7-10	7-11	7-11	8	8-1	8-1	8-2	8-3				
did	van	bad	red	cup	bee	lit	pin				62
8-3	8-4	8-4	8-5	8-6	8-6	8-7	8-8				
had	ran	pen	nut	big	old	yet	rob				70
8-8	8-9	8-9	8-10	8-11	9	9-1	9-2				
gun	leg	fun	lip	new	fog	has	sit				78
9-3	9-4	9-5	9-6	9-7	9-8	9-9	9-10				
sly	wig	mud	box	ink	sat	end	cut				86
9-11	10	10-1	10-2	10-3	10-4	10-5	10-7				
pay	fed	who	six	lad	wet	dry	cow				94
10-8	10-9	10-10	10-11	11	11-1	11-3	11-4	11-5			
his	peg	tin	say	eat	any	far	set	bud			103
11-7	11-8	11-9	11-11	12	12-1	12-3	12-4	12-6			
kid	pup	fox	ask	egg	cab	ill	use	jam			112
12-7	12-9	12-10	13	13-2	13-4	13-6	13-8	13-10			
all	pit	got	sad	tea	sky	one	yes	fur			121
14	14-4	14-8	15	15-6	16						
act	toe	her	our	ten	arm	rock	gone	feel			130
that	rich	till	long	flat	this	part	foot				138
made	upon	came	mile	back	sand	time					145
said	then	wall	into	were	done	walk					152
much	loss	seem	went	with	come						158

One-minute Reading Scale (Boys)

		5-6	5-7	5-7	5-8	5-8	5-9	5-10	5-10	5-11	
is	me	on	at	by	so	us	an	it	or	be	11
5-11	6	6-1	6-1	6-2	6-2	6-3	6-3	6-4	6-5	6-5	
to	as	he	of	in	go	up	am	if	no	we	22
6-6	6-7	6-7	6-8	6-8	6-9	6-10	6-10				
my	ox	do	the	and	for	but	him				30
6-11	6-11	7	7-1	7-1	7-2	7-2	7-3				
are	can	she	dog	let	you	not	was				38
7-3	7-4	7-5	7-5	7-6	7-7	7-7	7-8				
out	try	see	mix	cat	now	boy	saw				46
7-8	7-9	7-10	7-10	7-11	7-11	8	8-1				
bit	met	top	run	man	pet	lot	get				54
8-1	8-2	8-3	8-3	8-4	8-4	8-5	8-6				
did	van	bad	red	cup	bee	lit	pin				62
8-6	8-7	8-8	8-8	8-9	8-9	8-10	8-11				
had	ran	pen	nut	big	old	yet	rob				70
8-11	9	9-1	9-2	9-3	9-4	9-5	9-6				
gun	leg	fun	lip	new	fog	has	sit				78
9-7	9-7	9-8	9-9	9-10	9-11	10	10-1				
sly	wig	mud	box	ink	sat	end	cut				86
10-2	10-4	10-5	10-6	10-7	10-8	10-10	10-11				
pay	fed	who	six	lad	wet	dry	cow				94
11	11-1	11-3	11-4	11-6	11-7	11-9	11-10	12			
his	peg	tin	say	eat	any	far	set	bud			103
[illegible]	12-3	12-5	12-7	12-9	12-10	13	13-2	13-5			
kid	pup	fox	ask	egg	cab	ill	use	jam			112
13-7	13-10	14	14-3	14-6	14-9	15	15-4	15-8			
all	pit	got	sad	tea	sky	one	yes	fur			121
16											
act	toe	her	our	ten	arm	rock	gone	feel			130
that	rich	till	long	flat	this	part	foot				138
made	upon	came	mile	back	sand	time					145
said	then	wall	into	were	done	walk					152
much	loss	seem	went	with	come						158

One-minute Oral Addition Test

Question	Addition Age	
	Yrs.	Mths.
(1) 1 + 2		
(2) 4 + 1	5	0
(3) 2 + 2	5	3
(4) 2 + 4	5	6
(5) 3 + 2	5	9
(6) 4 + 3	6	0
(7) 2 + 5	6	3
(8) 5 + 4	6	6
(9) 3 + 5	6	9
(10) 8 + 2	7	0
(11) 4 + 4	7	3
(12) 5 + 2	7	6
(13) 6 + 4	7	9
(14) 1 + 8	8	0
(15) 3 + 7	8	3
(16) 6 + 3	8	6
(17) 2 + 6	8	9
(18) 5 + 5	9	0
(19) 7 + 2	9	3
(20) 4 + 6	9	6
(21) 7 + 5	9	9
(22) 8 + 3	10	0
(23) 4 + 9	10	3
(24) 6 + 8	10	6
(25) 7 + 6	10	9
(26) 9 + 8	11	0
(27) 9 + 6	11	3
(28) 8 + 7	11	6
(29) 5 + 9	11	9
(30) 7 + 9	12	0

One-minute Oral Subtraction Test

Question	Subtraction Age	
	Yrs.	Mths.
(1) 2 — 1		
(2) 3 — 2	5	5
(3) 5 — 1	5	9
(4) 6 — 2	6	0
(5) 5 — 3	6	3
(6) 2 — 2	6	7
(7) 7 — 2	6	10
(8) 6 — 4	7	2
(9) 7 — 3	7	5
(10) 6 — 3	7	9
(11) 8 — 2	8	0
(12) 7 — 5	8	3
(13) 8 — 3	8	7
(14) 7 — 4	8	10
(15) 9 — 3	9	2
(16) 8 — 5	9	5
(17) 10 — 4	9	9
(18) 9 — 5	10	0
(19) 10 — 3	10	3
(20) 9 — 4	10	7
(21) 11 — 2	10	10
(22) 10 — 6	11	2
(23) 12 — 3	11	5
(24) 11 — 6	11	9
(25) 12 — 5	12	0
(26) 13 — 4	12	3
(27) 15 — 9	12	7
(28) 14 — 6	12	10
(29) 17 — 8	13	2
(30) 16 — 7	13	5

Girls read a little faster than boys: hence the necessity for two scales. Let us suppose that a boy of 7 years and 3 months reads in the given time as far as the 43rd word ("cat"), without any mistake. His reading age is 7 years 6 months—the age found over the word "cat" in the scale for boys. If he makes three mistakes, we must go back three words and regard him as having read as far as "try." This makes his reading age 7 years and 4 months.

Now turn to the scale for girls. If a girl of 8 years and 2 months reads as far as "pet" (the 52nd word) and makes two mistakes we must count her as having read correctly up to "run." Her reading age would be 7 years 7 months; that is, she would be retarded by 7 months. If we wish to find her reading quotient we must divide her reading age (91 months) by her real age (98 months). This gives ·93 approximately. It is customary to multiply the result by 100 and to record the quotient as 93.

The standardisation of this test, which took place before the war, calls for a word of comment. The norms were secured by testing all the children from 6 to 10 years of age in a large number and variety of schools, and these norms were afterwards verified in the provinces. The scores for the individual schools did not vary by much more than six months on either side. The schools in a good neighbourhood were, as a rule, six months in advance of the normal, and the schools in poor areas six months behind. Although I tested some of the best schools in London I did not find any school which was a year in advance of the average. This was before the war; and before the introduction of individual methods of teaching reading. These

newer methods have rendered my norms too low. That is to say, if I restandardised the test this year I should probably have to raise the norms all round, and slightly lower the reading ages placed above the words. I have recently, for instance, found a school where the children are about a year and a half in advance of my norms. The reading ages given in the keys may be taken therefore to represent the average achievements under the older systems of teaching reading. This in no way detracts from the value of the scale as a means of gauging the efficiency of a special method of teaching reading, or as a means of securing a sound classification of scholars. It will, in any case, range children in the order of their reading ability.

There is yet another way in which the test can be used—a way which brings it more nearly within the category of intelligence tests. If the same children, exposed to the same teaching influence, are tested twice with an interval of three months, and graded on the ground of the progress made during those three months, the grading gives a rough index of their intelligence. For intelligence, as far as the school significance of the word is concerned, simply means capacity to learn.

Now for a few words of warning. The scale is of little use for older children. Ten years of age probably marks the limit of its usefulness. Moreover, it is of no use for anything else except testing. To use it for teaching is almost as bad as to measure a lad with a foot-rule several times a day with the idea of helping him to grow. A child learns to read fast by reading for pleasure, not by plodding through an arid list of words.

To compare the achievement of a whole school with the norms it is necessary to test all the children of a given age group. The children of exactly seven years of age in the top class (Standard I) of an infant school will nearly always have reading ages higher than seven. To find the real average for the school all the seven-year-old's in the lower classes must be tested as well.

The two oral arithmetic tests are given here in a form which will enable the examiner to find the arithmetical age of a child with the least possible trouble. Each child is, as in the reading test, examined individually and in isolation. He is asked the question, "One and two?" and as soon as he answers it he is asked the next, "Four and one?" and so on. He is not allowed to proceed until he has given the right answer. The examiner repeats the question, but gives the child no help of any kind.

The educational quotient of a child is obtained from these three tests in this way:

John Smith reads 59 words.	Reading age .	. 8 yrs. 4 mths.
„ „ adds 13 items.	Addition age .	. 7 yrs. 9 mths.
„ „ subtracts 12 items.	Subtraction age	8 yrs. 3 mths
	Total .	. 24 yrs. 4 mths.
	Average	. 8 yrs 1 mth.

$$\text{Educational quotient} = \frac{\text{Educational age}}{\text{Real age}} \times 100$$

$$= \frac{8 \text{ yrs. } 1 \text{ mth.}}{7 \text{ yrs. } 6 \text{ mths.}} \times 100 = 108.$$

John Smith's educational quotient is 108. It is this number which under normal scholastic condi-

tions and for the lower mental ages seems to accord closely with the Binet-Simon intelligence quotient.

What I have said about testing in the upper part of the infant school applies with equal force to testing in the lower part of the senior elementary school. This lower part is the part that really counts: it is the part that, from the point of view of good classification, demands from the head teacher both vigilance and insight. It is the placing of the children from seven to nine years of age that most clearly reveals the strength or the weakness of the school organisation. If the head teacher does not know these children through and through, if he does not know their more permanent mental qualities, if he does not know which are the runners, which the walkers, and which the crawlers, he is ignorant and neglectful of the resources of his school. And no amount of attention paid to the top classes will make up for that ignorance or remedy that neglect. The pace of mental growth during the first year in the senior school is faster than during any subsequent year. And the indifference sometimes shown to these younger children, these saplings, full of sturdy life and ready, under a little fostering care, to shoot out vigorously towards the light—the indifference sometimes shown them is amazing. They are put under the charge of the worst teacher in the school. They suddenly pass from the hands of the brightest teacher in the infant school to those of the dullest teacher in the senior school. And if perchance they happen to stray into the infant school to see their old teacher, they often astonish her by showing what little progress they have made. The opening buds that

left her full of promise now exhibit all the signs of having been chilled and checked by a sharp frost. And in the grip of that frost they are left to grow as best they can for at least a year. Sometimes for much longer. The serious thing is not so much that they are placed under the charge of a teacher without vision (this is sometimes inevitable), but that they are *left* under the charge of a teacher without vision. There is no one to observe them or to understand them. The mute inglorious Miltons remain mute, the Scotts sit indifferently on the dunce's stool, and the Darwins are treated as incorrigible dullards. This does not frequently happen; but it does sometimes happen. And it would never happen at all if the head teacher took care to supply what vision was lacking in the class teacher—if he constantly watched the younger children and frequently tested them. For if probing the mind is profitable anywhere it is profitable here. And pleasant, too. Children of this age like mental tests almost without exception. It is only the grown-ups who dislike being found out.

Let the tests be frequent, and let them be varied, and let them gauge the enduring things rather than the things that are easily gained and easily lost. The traditional mode of testing the children who come up from the infant school is not beyond criticism. To write a formal and cursive hand, to spell the commonest words, to work four sums on a piece of paper, one at each corner—these are trivial things compared with the capacity to read with fluency and to compute with ease and accuracy. At this stage it is what the child can do in his head that is important, not what he can put down on

paper. If he can put it down on paper as well, so much the better; but if he cannot do it now he can readily learn to do it later. The children in the infant school can easily be trained to work formal sums in the four fundamental rules, but they could spend their time much more profitably. While they are working four sums on paper they can work forty in their heads—smaller sums it is true; but these smaller sums are the bricks of which the larger sums are built; and even if we count the bricks the advantage is still overwhelmingly on the side of mental work. I submit, therefore, that the tests I have suggested for the infant school—oral and individual tests—the Binet tests and certain simple educational tests—are the best to apply to the children who enter the senior school from the infant school. There is no reason why a group test should not be applied as well, provided, of course, the children can read and write well enough to tackle it. And it is to meet some such need as this that I have devised the Group Test for Juniors, which appears in Chapter XX.

In recording and studying the results of testing a class or an age-group of children let us not forget the special merits of the median. The median represents the achievement of the group better than the average or arithmetical mean, because it is less disturbed by extreme or "freak" cases. It is not pulled down by a child at the bottom of the class who fails completely; nor is it pulled up by an exceptionally brilliant child at the top. On an adjacent page I try to exemplify the method and advantage of finding the median. The first list gives the marks obtained by a class of 31 boys

at a written examination. When the boys are arranged in order of merit it is found that the score obtained by the middle boy—the sixteenth boy—is 21. This score is the median. In a well-graded class the median is almost identical with the average. In this instance the average is 19·7, which is lower than the median because of the inordinate downward pull of the two bad papers at the bottom.

As the number of papers in the second batch is even there is no middle paper. The middle of the pack comes between the fourteenth and the fifteenth papers, and the median lies half-way between the marks obtained by these two papers. Both marks being the same the median is also the same, i.e. 35. The average, however, of all the marks on the list is 37·6, a level that is somewhat unfairly raised by the two abnormally high numbers at the top. It will be seen, therefore, that the median is an index of the true central tendency of a group: it points to the solid achievement of the majority of a class of children without being deflected by the achievements of a few members of exceptional brightness or exceptional dullness.

Assuming always that the papers are arranged in order of merit we take the middle one as the most significant; but there are two others which are almost as important. One is the middle paper of the top half, and the other the middle paper of the bottom half. The former is called the Upper Quartile, and the latter the Lower Quartile. In the first example the upper quartile is 24 and the lower quartile 15; in the second the upper quartile is 47 (half-way between 46 and 48), and the lower

MEDIANS AND QUARTILES

First Example

No.	Marks.	
(1)	30	
(2)	29	
(3)	28	
(4)	27	
(5)	26	
	26	
(7)	25	
(8)	24	← Upper Quartile
	24	
	24	
(11)	23	
(12)	22	
	22	
(14)	21	
	21	
	21	← Median
(17)	20	
	20	
(19)	19	
	19	
(21)	18	
(22)	17	
(23)	16	
(24)	15	← Lower Quartile
(25)	14	
	14	
(27)	12	
(28)	11	
(29)	8	
(30)	2	
(31)	0	

Second Example

No.	Marks.	
(1)	86	
(2)	70	
(3)	62	
(4)	59	
(5)	57	
(6)	51	
(7)	48	
		← Upper Quartile
(8)	46	
(9)	43	
(10)	41	
	41	
(12)	39	
(13)	36	
(14)	35	
		← Median
	35	
(16)	34	
(17)	32	
(18)	31	
(19)	28	
(20)	27	
(21)	25	
		← Lower Quartile
(22)	24	
(23)	22	
(24)	20	
(25)	17	
(26)	15	
	15	
(28)	14	

quartile 24½ (half-way between 24 and 25). From the quartiles we may gain a general notion of the way in which the marks are distributed. In the first example there is only a difference of 9 marks between the two quartiles; in the second example there is a difference of 22½. The second group is, therefore, more scattered than the first. It will be observed that the quartiles mark the limits of the middle half which constitutes the solid phalanx of a class.

It is by no means necessary to make out a sequential list of marks in order to discover the three significant levels of achievement. All we have to do is to see that the batch of papers is properly arranged. We take out the middle paper, which gives us the most important mark of all—the median. The median divides the batch into two equal groups. We find the medians of each of these smaller groups in precisely the same way as we find the median of the original group. These give us the two quartiles. Finally, we look at the first paper and the last; and these enable us to find the full range of marks. From these five papers alone a man may build up in his mind a fairly faithful picture of the whole batch.

Serviceable as are these five representative scores in recording the achievements of a group of pupils, or in comparing those achievements with standard scores or norms, we cannot dispense with the individual score of the individual child. For it is this that enables us to put the square peg in the square hole—to give the child his right teacher and his right syllabus of work. The marks obtained at any test, especially if it is an intelligence test, may

be so set forth as to provide a basis for searching criticism of the school organisation.

On page 159 I give a fictitious example of defective organisation in the lower part of a small elementary school for boys. The children are supposed to have been given an intelligence test, and the table is drawn up to record the age, standard, and score of each child. The second column shows the results for Standard I. Here a boy of seven got 50 marks, a boy of six got 47, and another boy of seven got 45. Two boys got 38 marks, one of them six years old and the other nine. The rest of the table should now be clear.

A cursory glance is sufficient to convince one that the general organisation is on the whole fairly sound. As the standards get higher the median age steadily increases, and so does the median score. But some of the extreme cases suggest serious misplacements. I say "suggests" rather than "proves," because there are other relevant facts which the Table does not disclose. The following criticisms, therefore, hold good only if the children's health is normal, and if tests of school attainments rank them in substantially the same order as the intelligence test.

Looking at the results for Standard I we note that the best boy, who is seven years of age, is far above the average for Standard III, and still further above the average for Standard II. In fact, there is only one boy in Standard III who can beat him. It is clear, therefore, that this Standard I lad is highly intelligent and ought to be in Standard III. It may be, of course, that he is a delicate child of great natural ability, but, through frequent absences

from school, of low scholastic attainments. This would account for his position, and would, so far, justify the organisation. He is, in any case, a boy to be watched with a view to rapid promotion. The same remarks apply to the next two pupils on the list. Both ought to be in a higher class—probably Standard III. The rest of the Standard I children are apparently in their right places.

The top boy in Standard II should be at least in Standard IV, and perhaps in Standard V. The boy of nine whose score is only 28 is probably over-promoted. He seems better suited for Standard I. Perhaps, however, he is a well-behaved and plodding boy who has earned his present position by sheer hard work. This can only be discovered by noting his position on the attainments list. The same kind of criticism applies to certain children in Standards III and IV. A new type of misplacement is exemplified by the boy of 12 in Standard III who gets only 27 marks, and the boy of 13 in Standard IV who gets only 31. These boys were obviously pushed up because of their age. The best thing to do with old and dull boys of this kind is to put them in a special class and set them to work on a special syllabus.

It will be seen that a table such as this is full of hints and suggestions. It points to possibilities and probabilities. It opens up lines of inquiry. And although in itself it settles nothing, in conjunction with other data it settles, or ought to settle, a great deal.

EXAMPLE OF DEFECTIVE ORGANISATION

Score	Standard I	Standard II	Standard III	Standard IV*
57				9
56				
55		8		12
54				11, 10, 10
53			10	10, 12
52				12
51				11, 10, 11, 12
50	7		9	10, 11
49				13
48			9, 11	
47	6			11
46			12	10
45	7		8, 9	
44			10, 9, 10	11
43		10, 9	9, 10	
42		8	8, 9, 11	
41		10	9, 10	
40		7, 9, 8		
39		8, 10, 8	10	
38	6, 9	9		
37		8	9	
36	6	10, 9	8	9
35	7, 8, 8	8, 8, 9		
34	7	8	10	
33	8, 6, 8	10		
32	7			
31	7, 7, 6			13
30				
29	8, 9			
28	9	9		
27			12	
26				
25				

CHAPTER XV

SILENT READING

WHEN the old examiner tests reading he asks the candidate to read aloud, and he records a score which serves as a reminder of how he likes the performance. The mark tells us as much about the examiner's taste in elocution as about the examinee's capacity to read. For, besides elocution, there is little else for the examiner to judge. All the older children read with a reasonable degree of fluency and accuracy, and if he is to differentiate them at all he must differentiate them mainly on the ground of those vague and elusive qualities, accent, intonation, and expression.

The new examiner tests oral reading, too; but only in the early stages. He measures, in a purely objective way, the degree to which the child has mastered the mechanical art of turning printed symbols into sounds. He does not attempt to assess either accent or expression: he confines himself to the assessing of accuracy and fluency. When, however, he examines a child who is old enough to understand what he reads the purpose of the inquiry is changed. Instead of gauging the rate at which the pupil can turn print into sound, he gauges the rate at which he can turn print into sense. In fact, he appraises the capacity of the pupil to profit by what he reads when he reads to himself.

It may briefly be said of the two types of examiner that one tests oral reading, the other silent reading; the old tests elocution, the new, comprehension. And, indeed, when we come to consider the proportion that silent reading bears to oral reading in

ordinary life, and when we remember that the only school pursuit which the pupils are quite certain to continue after they have left school is private reading, we see at once the importance of teaching and testing this pursuit with all the skill we can command. Private reading is not only a cardinal purpose in schooling; it has become for the pupil, under the present individualistic trend, a necessary means of progress. In educating a child by modern methods we assume a capacity to work alone and to extract knowledge from printed books. Professor McCall holds that if only one scholastic test is given in the school it should be a test in silent reading. And yet this is the one test which the old examiner never gives.

The aim of a silent reading test is to gauge comprehension; or rather the rate of comprehension —the speed at which the reader can absorb the right meaning from the printed page. There are several ways of doing this. One is to allow the child a fixed time for reading a piece of poetry or prose —generally prose—and then to require him to write out the substance of what he has read. This method is not fool-proof. It has all the mensural defects of the essay. Another way is to give the child a number of written or printed questions on the subject-matter, each of which may be answered by one or two words. A better method, perhaps, is the completion method which I adopted for the silent reading test given in *Mental Tests*. There the same passage is presented to the candidate a second time; but this time there are omitted certain key words which he is required to supply. The weakness of these methods is that they test two things—understanding and memory. The candidate must not only

understand what he reads at the time of reading, he must afterwards remember the detailed facts.

To test comprehension pure and simple the passage to be understood must remain before the pupil while he is being tested. He should have an opportunity of rereading it if he likes. It is this kind of test that is represented here. The test is similar in type to the paragraph-meaning test in the Stanford Achievement Examination, which is described in Chapter XIII. It is labelled B to distinguish it from my other silent reading test.

The examiner should make sure that the children know exactly what is required of them. After they have prepared their answer papers by inserting the numbers of the words so that nothing remains to be done but to write the words themselves, they should listen to a blackboard demonstration. The following paragraph should be written on the board.

Mr. Smith gave his son Tom a little dog and his daughter Jane a little cat. One day Jane's (51) began to run after Tom's (52).

The numbers 51 and 52 are suggested as there are no such numbers in the test. The examiner should use this example to explain and illustrate the instructions given at the beginning of the test paper.

The time allowed is 15 minutes; but if the pupils are over 14 years of age 10 minutes will be found a more appropriate period.

I have not yet applied this test to a sufficient number of children to be able to supply norms which could be said to represent the achievement of the "total population." So far as I can judge, however, the following represent the average for elementary school children:

Age:	9	10	11	12	13	14
Norms:	12	16	20	24	28	32

15 *Minutes*
10 *minutes for adults*

SILENT READING (B)

Wherever a number appears there is one word, and only one word missing. That word has to be written on your answer paper opposite its proper number.

Jane had two dolls, one with black hair and the other with brown. She liked the doll whose hair was brown, but did not like the doll whose hair was (1).

A little boy was sent to fetch three eggs from a nest and bring them to his mother. On the way back he fell down and broke two of them, so he was able to give only one (2) to his (3).

A man who was invited to dinner arrived at eight o'clock and found he was an hour late. He ought to have arrived at (4).

Anne was a very little girl who lived in the country and had to pass through a field and climb over a gate to get to school. One day she saw a big crow sitting on the gate. She was so frightened that she ran home to her mother. Now she never climbs the (5) without thinking of the (6).

George used to carry in his school bag, which had a small hole in it, a book, a pen, a pencil, and a piece of rubber. One day when he got to school he searched in his bag and found the book, the pencil, and the rubber, but he could not find the (7). He thought it must have dropped out through the (8).

A blind woman had a son who was the joy of her life. Her one fear was that he should become (9) like herself; and her one desire was that he should (10) her as much as she loved him.

There was once a parrot that had learnt to speak, but could only say two words, "Yes" and "No." He always, however, used them wrongly, and the answer he gave was always the one that was not expected One day a dear kind old lady gave him a piece of sugar and asked: "Do you like me?" The (11) at once answered (12).

A man stumbled into a gutter in which a pig was lying. The man had a ring on his finger, the pig had a ring in his nose. The man was drunk, the (13) was sober. It sometimes happens that a (14) is a more attractive creature than a (15).

To make a good cup of tea it is necessary that the teapot should be heated before the boiling water is poured in, and that the boiling water poured in should not have been left boiling for a long time. One day Martha undertook to make tea for the family. She put the right quantity of tea in the pot and poured on it some water which had only just (16). But the tea was not so good as it might have been, because she had forgotten to (17) the (18).

Mr. Robins liked to ride outside the bus when the weather was fine, and inside when it was wet. But he generally found to his disgust that when the day was fine he could not ride (19) because it was crowded; and when the day was (20) he could not, for the same reason, ride (21).

A boy called Henry, who was born in France, used to attend an English school with his little sister. He spoke English with a (22) accent and was always quarrelling with boys who he thought were unkind to his (23).

A lady once wrote to Mr. Watson, her bookseller, asking him to send her certain books which she named. She also wrote to Mr. Walton, her fishmonger, ordering three pounds of salmon. The names were so similar that she confused them and put the letters in the wrong envelopes. When Mr. Walton got his letter he was much surprised to find himself asked to supply (24); and when Mr. (25) got his, he muttered: "Does the lady take me for a (26)?"

In the branches of a great tree lived a wise old crow. His wife was dead, and his children were getting their own living; so he had nobody to look after but (27). He mourned the loss of his (28), and when he visited his (29) he often used to talk to them about their (30).

I did not like the sermon to-day. In the first place the preacher read his sermon; in the second place he did not read it well; and in the third place it was not worth (31).

Sugar did not come into use until about the middle of the Middle Ages. Before that time honey was probably the sweetest thing known. When the ancient Greeks wanted to make their food (32) they used to add (33).

If a substance is lighter bulk for bulk than the liquid in which it is placed it will float in that liquid; if it is heavier it will sink. Ice being (34) than water floats in it; and when a piece of iron is thrown into a bath of quicksilver it will (35) because quicksilver is heavier than iron.

It is a mistake to think that bright flowers flourish most luxuriantly in the tropics. It is rather in the cold and less favoured regions of the world that one must look for fine floral displays and bright masses of colour. When, therefore, we climb high mountains and approach the snow-line, we find the (36) more and more abundant, and their (37) more and more brilliant.

Society, we believe, is constantly advancing in knowledge. The tail is now where the head was some time ago. But the head and the tail still keep their distance. A labourer to-day reads better than a nobleman of the Middle Ages; but a (38) of to-day (39) better than a (40) of to-day.

Whenever all the world around is remarkably uniform in colour and appearance, all the animals, birds, and insects alike disguise themselves in its prevailing tint to escape observation. Animals in the desert take the colour of the sand. In the Arctic snows all animals, without exception, are (41). In the (42) they are yellow.

Trees receive their nourishment from two sources, the soil and the air. By means of their roots they suck up water with minerals dissolved in it; and by means of their leaves they absorb moisture and gases from the air. One of those gases is their most important food, for it supplies them with carbon, of which nearly the whole of their solid substance is composed. It is, in fact, the lesser part of the nourishment of the (43) that comes from the (44); the greater part is taken in through the (45).

Perhaps the most curious instance of absence of mind is that of a gentleman who after taking his bath one morning dried himself with his newspaper, and sat down to (46) his (47).

An Englishman and a Frenchman quarrelled, and agreed to fight a duel in a dark room. The Englishman, not wishing to have murder on his conscience, groped his way to the fireplace, fired up the (48) and brought down the (49). In the French version of this story it is the (50) who is made to hide up the chimney.

KEY TO SILENT READING TEST

1. black.
2. egg.
3. mother.
4. seven.
5. gate.
6. crow
7. pen
8. hole.
9. blind.
10. love.
11. parrot, bird.
12. no.
13. pig.
14. pig.
15. man.
16. boiled.
17. heat, warm.
18. pot, teapot.
19. outside.
20. wet.
21. inside.
22. French, foreign.
23. sister.
24. books.
25. Watson.
26. fishmonger.
27. himself.
28. wife.
29. children.
30. mother.
31. reading, hearing.
32. sweet.
33. honey.
34. lighter.
35. float.
36. flowers, blooms, blossoms.
37. colour(s), hue(s), tint(s).
38. nobleman.
39. reads.
40. labourer.
41. white.
42. desert.
43. tree.
44. soil, ground.
45. leaves.
46. read.
47. towel.
48. chimney.
49. Frenchman.
50. Englishman.

CHAPTER XVI

ENGLISH

An English examination without an essay seems very much like the play of *Hamlet* without the Prince of Denmark. And yet it is highly probable that a new examination in English can test nearly all the qualities which an essay is specially supposed to exhibit, and test them with a reasonable degree of certitude. Nearly all, but not quite; for to deal adequately with the power to initiate and organise ideas is a task which eludes the ingenuity of the new examiner. But, then, it also eludes the ingenuity of the old examiner. The old examiner tries to evoke a manifestation of creative power, and perhaps succeeds in evoking it. But he cannot measure it: he can only estimate it by guesswork. Under the old system the examiner was allowed to guess, but the candidate was not; under the new system the candidate sometimes guesses, but the examiner never.

Of the two examinations here given the first has no claims to originality; except perhaps in applying the completion method to the testing of punctuation, and of those errors of composition which a recent research of mine has shown to be most prevalent.[1] The examination as a whole gauges the candidate's familiarity with, and understanding of, the mother tongue. A fair proportion of the items bear on those formal matters which are sometimes called "the mechanics of English." The second examination is quite another matter. It is

[1] See *Group Tests of Intelligence*, pp. 246–9.

an attempt—the only attempt of its kind so far as I am aware—to supply one of the functions of the missing essay. The candidate is required to organise his ideas just as he has to organise them in writing an essay. The elements, however, instead of being selected by him are selected for him. What he is ostensibly asked to do is to arrange phrases and sentences: what he is really asked to do is to arrange the underlying thoughts.

This constructive test is an extension of the dissected sentence test used by Binet and adopted by the examiners for the American Army. There separate words have to be rearranged to form a sentence; here separate phrases and clauses have to be rearranged to form a sentence, and separate sentences have to be rearranged to form a paragraph. In the first seven paragraphs the elements are phrases and clauses; in the rest of the paper the elements are sentences. The passages given were extracted, sometimes with a trifling alteration, from the writings of Thackeray, Kenneth Graham, Aldous Huxley, Walter Bagehot, Augustine Birrell, Prentice Mulford, Olive Schreiner, and Laura Richards. Most of these authors were chosen because of their fondness for the short sentence. Long sentences would perhaps do as well, but they would occupy more space, cost more for printing, take a longer time to read, and show, as a rule, a more obvious and inevitable sequence.

The tendency of modern English, as compared with the English of the nineteenth century, is towards the short sentence and the omission of connectives. We do not now—not at any rate if we wish to avoid the charge of pedantry and affectation—build our sentences with great elabora-

tion, with great care, and with due regard for logical co-ordination and subordination; imitating, as did Ruskin in his early years, and many another ambitious writer in the more leisured days of the eighteenth century, the long, involved periods of Hooker; rearing an intricate structure of words, phrases, and clauses, all linked together and dependent on one another as the parts of a Chinese puzzle are dependent on one another, and making the same impression on the mind as a group of acrobats who, in their final and crowning feat, just before the curtain falls, stand on each other's shoulders and on each other's hips, and hang in extended order from each other's arms so as to form an imposing pattern of human limbs, trunks, and heads. We have ceased to do this. We abbreviate our sentences. We emancipate most of the dependent clauses and give them the full franchise of the sentence. We drop most of our "since's" and "because's" and "therefore's," and show no great love for the causal and illative tribe of conjunctions. But when we do not use these words we think them. The mind supplies the necessary relation, and in supplying the necessary relation it supplies the necessary cohesion. And as in the simple sentence the disappearance of inflexions has made the order of words of primary importance, so does the disappearance of connectives between the larger units of speech make the order of those larger units of primary importance.

Although I have called the tendency to use the short sentence a modern tendency, there is nothing new in the style. Jingle had it as a disease. Dr. Hugh Blair, in his *Lectures on Rhetoric*, published

in the eighteenth century, refers to it as the *style coupé*, and quotes the following example from Pope: "I confess, it was want of consideration that made me an author. I writ, because it amused me. I corrected, because it was as pleasant to me to correct as to write. I published, because, I was told, I might please such as it was a credit to please." If we shuffle the sentences in a passage written in *le style coupé* and require their rearrangement in logical sequence we set a task which calls into play not so much the candidate's knowledge of the clothing of ideas as of the anatomy of ideas. He has to find the bones of the paragraph and articulate them aright.

The reader who himself attempts the test in Constructive English will soon discover that the solution of most of the items is obvious and inevitable: there is only one order that will make sense. There are, however, one or two of the fifteen paragraphs where the solution is not so simple—where it is not a question of finding the one and only possible order but of finding the best among a number of possible orders. The solution I give in the key represents the order adopted by the original authors; and if the reader will compare that with any other arrangement that seems to him logical and coherent he will, I think, admit that there are sound reasons for preferring the original. The points of superiority are, however, few and inconspicuous; and I must confess that when I attempted, after a fairly long interval, to work the paper I felt glad that I had preserved the key. This does not, I venture to think, detract from the value of the test: it merely means that the bulk of the marks can be got on a sequence which is fixed by the laws

of logic and common sense, and the rest of the marks on a sequence which conforms to the more flexible canons of propriety and taste.

The scoring of the paper is not easy. A mark is given for each single sequence that is correct, the first being regarded as the sequence 01. In other words, a mark is given if the first ordinal is right, and a mark for every other ordinal that follows its right antecedent. We will suppose, for instance, that a group of pupils have in front of them worked papers ready to be marked. The examiner instructs them to cross out every number that does not follow properly, and to leave alone every number that does. Then he starts dictating the answer to the first item (2143) in this fashion: 1 follows 2; 2 comes first; 3 follows 4; 4 follows 1. The score for that item is found by counting the ordinals not crossed out. For the whole paper the maximum score will be seen to be 100.

I have not applied these tests very extensively, but I have sufficient data to put forward the following scores as tentative norms for elementary school children:

ENGLISH (COMPREHENSION)

Age	10	11	12	13	14
Norms	30	36	42	48	54

The best school tested was two years in advance of the normal, and the worst two years behind. Central schools are at least three years in advance.

ENGLISH (CONSTRUCTION)

Age	10	11	12	13	14
Norms	14	20	26	32	38

In this test the central schools are at least four years in advance.

One Hour

ENGLISH (COMPREHENSION)

Choose from among the four words in brackets the word that means the opposite of the first word.

Sample : Dry (cold, windy, wet, hard). Answer —Wet.

1. Cautious (guarded, adverse, harsh, rash).
2. Sympathy (antipathy, passion, disrespect, courtesy).
3. Purpose (permit, chance, project, stimulus).
4. Wisdom (sagacity, folly, sanity, jargon).
5. Remembrance (surprise, imagination, death, oblivion).
6. Certain (probable, impossible, doubtful, credible).
7. Offer (refuse, present, request, protest).
8. Severe (unruly, faithful, angry, lenient).
9. Permit (tolerant, prohibit, consign, consent).
10. Cunning (skilful, truthful, pleasant, artless).

When two words mean almost the same write S for same.
When two words mean almost the opposite write O for opposite.
When you do not know which they are write N.K. for not known

Sample : Quick	Rapid	Answer—S.
11. Before	After.	
12. Hide	Conceal.	
13. Candid	Frank.	
14. Public	Private.	
15. Affirm	Deny.	
16. Puncture	Perforation.	
17. Shallow	Profound.	
18. Irritable	Imperturbable.	
19. Observance	Compliance.	
20. Agreement	Compact.	

From among the words or stops in brackets select the word or stop that makes the best sense. The number of the question is put before the bracketed words.

Sample: She invited me 15 (for to after) tea. Answer: 15 to.
What a pity 16 (. ? !). Answer: 16 !.

I do not wish 21 (of by for) more money.

The sun rises 22 (earlier later sooner) in summer 23 (but though than) in winter.

24 (Their There It) is no disgrace 25 (on with in) poverty.

How far is it from here to Scotland 26 (. ? ;).

He asked how many 27 (weeks weaks) there 28 (where wear were) in a month 29 (, ? .).

Water is good 30 (, . !) but milk is better.

31 (To Two Too) miles is 32 (to two too) far 33 (to two too) walk 34 (to two too) school.

They 35 (have do does) not wish 36 (their our there) names to be known.

The men 37 (did done) their duty, and the women have 38 (did done) theirs.

Everybody likes to see 39 (their his one's) name in the newspaper.

40 (Whom Who That) did you see there ?

Mr. Jones, 41 (who which whom) I knew well, was elected chairman.

Mr. Jones, 42 (who whom what) I knew to be a good man, was elected chairman.

Mr. Jones, 43 (who whom which) I know is a good man, was elected chairman.

They found no fault with my wife and 44 (I me we).

They believed that 45 (us we) two had done it.

They saw 46 (us we) two crossing the street.

He 47 (as has) to eat 48 (is his) dinner 49 (as has) soon 50 (as has) it 51 (is his) ready.

Paul 52 (come came) to meet his father as soon as he 53 (saw seen) him.

54 (Sarahs Sarah's) cat had 55 (its it's) fur singed.

Both they and we grow potatoes, but 56 (theirs their's) are larger than 57 (ours our's).

" We must burn them, I think, 58 (? " —) said Miss Matty, 59 (thinking laughing looking) doubtfully at me 60 (. " ,). " No one will care for them 61 (where when why) I am gone " 62 (; . !). And one by 63 (one two three) she dropped them into the 64 (side top middle) of the fire, 65 (feeling looking watching) each blaze up, 66 (die live breathe) out, and rise away up the 67 (grate, chimney, mantel-piece) before she gave 68 (one some another) to the same fate.

The 69 (shouting cackling cawing) of the rooks in February shows that the 70 (time year incident) is coming when their 71 (houses nests lairs) will be 72 (occupied reoccupied destroyed) 73 (, ; .).

They 74 (jump resort walk) to the trees, and 75 (perch squat quarrel) above the old nests to 76 (mark sing indicate) their 77 (rights wrongs pleasures); for in the old 78 (barn trees rookery) possession is the law, and not 79 (nine-tenths, four-fifths, two-thirds) of it 80 (entirely, only, absolutely).

There is no place in the 81 (country town continent) which I love so much to 82 (frequent describe emphasise) as the Royal Exchange. It gives me a secret 83 (annoyance jest satisfaction), and, in some 84 (tone measure sense) gratifies my 85 (vanity mind uncle), as I am an Englishman, to see so rich an assembly of countrymen and 86 (friends, foes, foreigners) consulting 87 (apart together often) upon the private 88 (business quarrels, politics) of mankind and making this 89 (world metropolis country) a kind of emporium for the whole 90 (town island earth).

Name by letter (A, B, C, etc., below) the sentences that best give the meaning of each of the following proverbs.

Sample: Still waters run deep. Answer: G.

91. A stitch in time saves nine.
92. We catch more flies with honey than with vinegar.
93. Birds of a feather flock together.
94. Those who live in glass houses should not throw stones.
95. One man's meat is another man's poison.
96. Out of the frying-pan into the fire.
97. A rolling stone gathers no moss.
98. One swallow does not make a summer.
99. Penny wise, pound foolish.
100. Once bit, twice shy.

A. A man is known by the company he keeps.
B. A change for the worse.
C. Experience doth make cowards of us all.
D. Don't change your occupation too frequently.
E. Delays are dangerous.
F. Careful over little things and careless over big.
G. Silent people are the most thoughtful.
H. Kindness is more effective than harshness.
I. Different things suit different people.
J. Don't judge on too slight evidence.
K. Only the faultless should find fault.

KEY TO ENGLISH (COMPREHENSION) TEST

1. Rash
2. Antipathy
3. Chance
4. Folly
5. Oblivion
6. Doubtful
7. Refuse
8. Lenient
9. Prohibit
10. Artless
11. O
12. S
13. S
14. O
15. O
16. S.
17. O
18. O
19. S
20. S
21. for
22. earlier
23. than
24. There
25. in
26. ?
27. weeks
28. were
29. .
30. ,
31. Two
32. too
33. to
34. to
35. do
36. their
37. did
38. done
39. his
40. Whom
41. whom
42. whom
43. who
44. me
45. we
46. us
47. has
48. his
49. as
50. as
51. is
52. came
53. saw
54. Sarah's
55. its
56. theirs
57. ours
58. "
59. looking
60. .
61. when
62. .
63. one
64. middle
65. watching
66. die
67. chimney
68. another
69. cawing
70. time
71. nests
72. reoccupied
73. .
74. resort
75. perch
76. indicate
77. rights
78. rookery
79. nine-tenths
80. only
81. town
82. frequent
83. satisfaction
84. measure
85. vanity
86. foreigners
87. together
88. business
89. metropolis
90. earth
91. E
92. H
93. A
94. K
95. I
96. B
97. D
98. J
99. F
100. C

One Hour
Scrap paper allowed

ENGLISH (CONSTRUCTION)

Each of the sentences or paragraphs in this test has been disarranged by moving words, or groups of words, from one place to another. Your task is to rearrange them so as to make the best sense. Each group of words is preceded by a number, and the answer you are required to give should consist of these numbers in their proper order.

For example:

1 till he is ill 2 should fast 3 the boy 4 till he is well 5 who eats.

As it stands this is nonsense; but if you read it in the order 35124 it makes good sense.

The answer therefore is 35124.

Again, look at this:

1 As he was very little his mother tied him to the string of her apron. 2 Once upon a time a boy played about the house. 3 "Now," she said, "you cannot stumble."

The correct order to this is 213.

Rearrange the following in the same way:

(1) 1 people do not know 2 it is a pity 3 think of them 4 what other people.

(2) 1 and the battle-axe the only argument 2 in the ancient state of society 3 force the only teacher 4 war was the only trade.

(3) 1 or other piece of crockery 2 as for me 3 I touch a cup 4 I can only say 5 I seem to upset it 6 that every time.

(4) 1 from school 2 which Amelia had sent 3 and leave the room 4 once 5 which caused her to burst into tears 6 Rebecc suddenly came across one 7 in looking over some drawings.

(5) 1 which formed the usual conclusion 2 at Mr. Osborne's house 3 and departed 4 and drunk the glass of wine 5 as soon as the young ladies had eaten an orange 6 of the dismal banquets 7 to make sail for the drawing-room 8 the signal was given 9 and they all rose.

(6) 1 took a most unfairly long pull 2 who had evidently waited 3 announced 4 for a walk 5 and then jumping up 6 a third bottle was by this time circulating 7 for it to reach her 8 and Selina 9 and shaking out her frock 10 that she was going.

(7) 1 and therefore passed over 2 that even the Mole 3 "You've been a bit of a duffer this time, Ratty!" 4 the Badger's caustic 5 as far as possible 6 not to say brutal 7 could not help saying 8 remarks may be imagined 9 though he took his friend's side 10 but it was painful to the Rat.

(8) 1 All the trains—the few that there were—stopped at all the stations. 2 Along this particular stretch of line no express ever passed. 3 Denis knew the names of those stations by heart.

(9) 1 But Falstaff constantly tells us about his earlier life. 2 There is nothing, or next to nothing, in Shakespeare's works which throws light on his own story. 3 The difficulties of a biographer are, however, different in the two cases. 4 There is more material for a life of Falstaff than for a life of Shakespeare.

(10) 1 There is little trace of labour in his composition. 2 You seem to read of good wine and good cheer. 3 It is poured forth like an unceasing torrent. 4 His words have a flow, a vigour, which is not given to hungry mortals. 5 Sydney Smith was an after-dinner writer.

(11) 1 That is not true. 2 Our graveyards are full of lies. 3 Our friend is not dead. 4 We place a stone over the cast-off body of a friend. 5 It is only the body he used that lies there. 6 We place on that stone the word "died."

(12) 1 A blind woman had a son, who was the joy of her life. 2. Moreover they are baby clothes, and my play-fellows mock and laugh at me because of them." 3 By and by the boy came to her and said: "Mother, give me some other clothes to wear.

4 Though she had no sight in her eyes yet she was skilful of her hands; and it was her delight to make pretty clothes for her boy. 5 See how soft and warm they are." 6 But the mother said: "Nay, my darling: these are by far the best clothes for you. 7 These are too small for me; they pinch and bind me.

(13) 1 But there is another method—the method of the life we all lead. 2 There is a sense of satisfaction in all this, and of completeness. 3 Human life may be painted according to two methods. 4 Men appear, act, and react upon each other, and pass away never to return. 5 According to that each character is duly marshalled at first and ticketed; we know with an immutable certainty that at the right crisis each one will reappear and act his part. 6 One is the stage method. 7 Here nothing can be prophesied.

(14) 1 The front door stood hospitably open. 2 He loosed his grip of the brakes, and in a moment was rushing headlong down. 3. There was nobody to take. 4 Five minutes later he was rushing through the gate of the courtyard. 5 The place was entirely deserted. 6 The hill down which he cycled was becoming steeper and steeper. 7 But he took nobody by surprise. 8 He would take them by surprise. 9 He left his bicycle leaning against the wall and walked in.

(15) 1 "You may not know him, but he will know you." 2 "I have no brother," said the child. 3 "This is really shocking. 4 One day the Tidy Angel came into the nursery. 5 There was once a child who was untidy. 6 "Yes, you have," said the Angel. 7 You must go out and stay with your brother while I set things right here." 8 "This will never do," said the angel. 9 He left his books on the floor and his muddy boots on the table.

KEY TO ENGLISH (CONSTRUCTION) TEST

(1) 2143
(2) 2431
(3) 246315
(4) 4721653
(5) 541628793
(6) 68271593[10]4
(7) 4681[10]29573
(8) 213
(9) 4321
(10) 54213
(11) 246135
(12) 1437265
(13) 3652174
(14) 624198735
(15) 594837261

CHAPTER XVII

MATHEMATICS

IT is currently believed that the easiest subject to examine is Mathematics. Not because a mathematical paper is easy to set, but because it is easy to mark—easy and satisfying. The examiner feels much more confident in the fairness of the final score than when he marks a paper in English, in History, or in Natural Science. The scale being objective and the marking mechanical there is no room for individual judgment. All examiners working on the same plan inevitably make the same appraisal.

But this optimistic view can only be maintained if the answers are marked in a simple scheme of "right or wrong." If in an arithmetic paper a candidate secures full marks when a sum is right and no marks at all when it is wrong, then complete objectivity is obtained, it is true, but at too great a cost. The examination has become "fool-proof" at the cost of becoming unjust. Under the conditions prescribed it is possible for a candidate to be quite a good arithmetician, and even to show ample evidence of it in his paper, and yet secure no marks at all. A slight slip in calculation in any one of the answers would not only cancel the credit earned

by a large number of correct computations, but would invalidate the result of a perfect piece of mathematical reasoning.

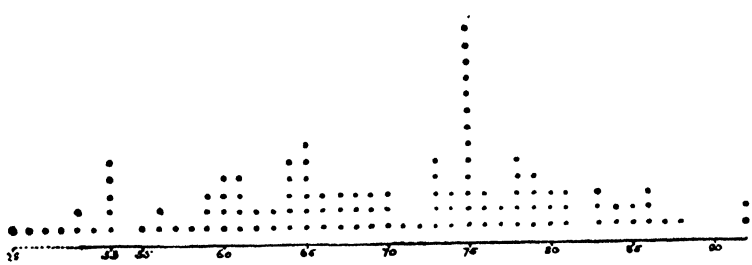
It may, however, be contended that mathematical papers are not marked in that way. Credit is always given for answers that are partly correct. This is quite true; but once we depart from a rigid right-wrong system of marking we open the floodgates of personal vagaries and lose the benefit of a scale which is objective and invariable. Some years ago a preliminary examination for scholarships was held simultaneously in a large number of elementary schools, and the papers were marked by the head teachers. One of the arithmetic questions ran as follows:

> There are twice as many books on this side of the table as on that side. There are three more books on this side than on that. How many books are there on the table?

Many of the candidates gave the correct answer and nothing else: they made no attempt to show the working. Some of the head teachers argued that in a case like this the result might have been arrived at by mere guessing, and that there was no evidence of a logical process of thought having been followed. They consequently scored the result nought. Others urged that the candidate was able to think out the solution even though he could not clearly formulate the steps. They accordingly gave half-marks. Others again maintained that work done in the head was a better index of ability than work done on paper; and these not only gave full marks, but would have given more if more were

permissible. This diversity of practice might, of course, have been avoided by definite instructions issued to all the examiners. But however detailed and comprehensive the instructions they can never meet the innumerable possibilities of error in a complex mathematical example. The examiner has at many points ultimately to rely on his own private judgment.

In 1913 Dr. Starch and Dr. Elliott made an investigation in which a paper worked at a final examination in Geometry at an American high school was submitted to 115 mathematical teachers, all of whom marked the paper independently. The distribution of the several ratings is shown in the accompanying diagram.



Distribution of marks assigned by 115 mathematics teachers to a final examination paper in Geometry. After Starch and Elliott.

It will be seen from the figure that the marks ranged from 28 to 92. The conclusion reached by Dr. Starch is that "mathematical papers are not marked with mathematical precision any more than other papers are." Quite clearly the examiners in this case marked on different principles. Each had his own scheme of scoring, each his own scale of

values. If all these examiners had been brought together to discuss the basis on which marks should be awarded and to agree upon a common plan of procedure the range of variability would be narrowed. But the variability would still be there.

We conclude, therefore, that a mathematics examination as at present conducted cannot lead to an objective score—a score independent of the personality of the examiner—except by a very crude and unjust scheme of marking; and that once a reasonable measure of refinement is introduced the result has much of the uncertainty and precariousness which pertain to the marking of essays. It is essential in good testing that the scale of measurement should be objective and not subjective, and that it should be fine and not crude. But in mathematical examinations when the scale is objective it is crude; and when it is fine it is subjective.

The obvious remedy is to refine upon the objective scale. The objective scale is crude, not because it is objective, but because the unit of measurement is absurdly large. And this large unit measures an extremely complex product. For the unit is one question, which takes ten or fifteen minutes to answer, and which involves in its solution many processes and many operations. In the whole-or-none method of marking there is nothing wrong; indeed, it is the only possible method of marking. It is the method that everybody in the last resort adopts whether he knows it or not. Let us suppose, for instance, that he has to mark a sum in which some of the working is right, though the final answer is wrong. He immediately begins to analyse

the work, and, taking each section or aspect by itself, gives full credit for each part that is right and no credit for any part that is wrong. It is the right-wrong method applied to the details instead of the whole. And the variability in the scores arrived at by this method is not due to the method itself, but to imperfect analysis of the answer and to difference in the evaluation of the several parts.

And since analysis is necessary, why should it not be made once for all by the deviser of the test instead of being made a hundred times over by the markers of the test? Why not, in other words, set a large number of small sums instead of a small number of large sums? This is precisely what the new examiner does. At the London University matriculation, which typifies the examination of yesterday and to-day, the arithmetic and algebra paper consists of ten questions which have to be worked in three hours. The arithmetic paper used at the American Army Intelligence Examination, which, perhaps, forestalls the examination of to-morrow, contained 20 questions which have to be worked in 5 minutes. While in one examination a question takes 18 minutes to answer, in the other it takes 15 seconds. The matriculation sums are long and complex; the army sums are short and simple. The matriculation marker (or "reader," as he is often called) has to evaluate partial results; the army marker has merely to count each sum that is right. The former is a reputable mathematician; the latter a clerk who works with a stencil.

The question arises: Is the new type of examination question any better than the old? Is it,

indeed, as good? Is it conceivable that simple questions in mental arithmetic, however numerous and varied they may be, can compare in diagnostic value with the difficult questions set at a university examination? And if not, do we not pay too high a price for our fool-proof test? Rather than discuss these questions on *à priori* grounds, let me give an account of an investigation I made with a view to discovering the relative merits of the old system and the new.

To represent the new system I got out the two examination papers, one in mechanical arithmetic and the other in arithmetical reasoning, papers which will be found on pages 190–198. The pupils were first instructed to prepare their answer papers by inserting name, age, and standard, and by putting down the numbers of questions from 1 to 100; so that nothing remained to be written except the answers. The question papers were then distributed, and 50 minutes allowed for the first paper and 60 for the second. The two papers were quite distinct —a day's interval was sometimes allowed to elapse between them—but the results were combined. It was really one examination in two parts. The children wrote with lead pencil instead of ink, and were permitted to use the back of the answer paper for scrap work.

The first investigation was made at a large central school for boys. The fourth-year boys (76 in all) were chosen for the experiment. The marks obtained by each of these boys at all the terminal arithmetic examinations held during the previous three years were added together and a list drawn up in an order of merit based on the aggregate marks.

If any boy had missed an examination he was credited with the average of the marks gained at the preceding and at the succeeding examination. The composite list thus arrived at was regarded as a standard record of the relative arithmetical abilities of the boys. There was, I assumed, greater stability and trustworthiness in the order of merit thus obtained than in an order based on achievement at any one examination, however searching that examination might be. At any rate, this final ranking was taken as the criterion or standard by which I estimated the relative values of the two types of examination. The boys were now given two examinations, one of the old type and one of the new.

The old type was represented by an examination which had been set some years previously for L.C.C. Trade Scholarships. The paper, which was quite difficult enough to give a reasonable distribution of marks, consisted of ten questions which had to be answered in 1 hour and 20 minutes.

The new type was represented by the two papers already described.

Thus, in addition to the standard order, I obtained two others, one based on the old examination and the other based on the new; and each of these new orders was compared by the standard order. They were compared with the method of correlation, and the co-efficients calculated by the Spearman formula for ranks. The outcome was that the trade scholarship test showed a correlation of ·36 (*p.e.* = ± ·07), while my tests showed a correlation of ·76 (*p.e.* ± ·03). In other words, the order obtained by the new method of testing tallied much

more closely with the standard order than did the order given by an examination of the traditional type.

My next step was to confirm this conclusion by varying the conditions of the experiment. The top class of an ordinary elementary school was chosen. It consisted of 38 boys. They were given a searching examination in arithmetic on each of ten successive school days. These examinations were made so varied and comprehensive as to cover every kind of example which the boys were capable of working. The marks were totalled and an order of merit arrived at which was used as the criterion for comparison. The old examination was in this instance exemplified by a paper set some years before for the L.C.C. Junior County Scholarships. The new type of examination was represented by my two papers as before.

The correlations were higher all round. The correlation of the Junior County Scholarship test with the standard was ·82 ($p.e \pm$ ·03), that of my tests with the standard was ·9 ($p.e. = \pm$ ·02). When each of my tests was taken separately, however, the correlation was not so high. It was ·81 for the mechanical test and ·79 for the problems. We therefore get a more satisfactory finding if we use both tests than if we use one—a conclusion which the reader will, no doubt, have reached without the aid of statistics.

Once again was the experiment made, in another school. Ten examinations were again given; and the result of my tests was found to correlate more highly with the summarised result than did that of the most difficult of the ten separate examinations.

We seem, therefore, to have good experimental ground for believing that the new examination is better than the old. As the two new arithmetic tests have been given to about 15,000 children in elementary schools, they may be regarded as fairly well standardised. There is a marked difference in the achievements of the two sexes, and a marked similarity in the results for the two tests. The norms giving the average number of sums right by age-groups and by standards are as follows :

MECHANICAL ARITHMETIC

Age:	9	10	11	12	13	14
Norms (boys)	11	20	29	38	46	53
Norms (girls)	11	17	23	29	36	44

ARITHMETICAL REASONING

Age:	9	10	11	12	13	14
Norms (boys)	12	23	35	41	47	53
Norms (girls)	10	15	20	31	37	43

MECHANICAL ARITHMETIC

Standard	II	III	IV	V	VI	VII	Ex-VII
Norms (boys)	6	13	21	32	44	58	73
Norms (girls)	6	10	18	26	39	52	62

ARITHMETICAL REASONING

Standard	II	III	IV	V	VI	VII	Ex-VII
Norms (boys)	6	13	21	32	44	58	73
Norms (girls)	4	9	17	25	38	51	60

The best school tested was nearly two years in advance of these norms, and the worst school

nearly two years behind. This accords fairly well with the difference in mental ages between these two extreme schools.

Care should be taken to interpret the age aright. The norm given for 9 years of age refers to children who are precisely 9 years old and not to the group of children whose ages range from 9 years to 10 years and whose average is 9 years and 6 months. The age-group included in the term "9-year old's" embraces children from 8½ to 9½; they are the children whose nearest birthday is 9.

In comparing the scores for a school with the norms care should be taken to include all the children in the school who are of the given age. To exclude those in the lower standards will falsify the comparison.

In central schools the norms are about three years in advance of those at ordinary schools.

The algebra test has been set in two central schools for boys and the averages secured are given below. The "years" refer to the time during which algebra has been studied:

Years . . .	1	2	3	4
Norms . . .	33	37	56	69

It is difficult to account for the irregularity in the progress made in successive years.

Fifty Minutes

MECHANICAL ARITHMETIC

Work in your head if possible. Scrap paper allowed.

1. $8 + 5$.
2. $9 + 6 + 8$.
3. $8 + 7 + 13$.
4. $34 - 19$.
5. $8 + 9 - 3 + 2$.
6. $17 + 18 - 14$.
7. 3,000 — 1.
8. $156 - 156$.
9. $200 - 13$.
10. $54 + 368 + 7$.
11. $100 - 4 - 96$.
12. $201 - 102$.
13. 47×8.
14. $63 \times 5 \div 9$.
15. 587×7.
16. $1 \div 1$.
17. $5 \div 1$.
18. $0 \div 4$.
19. $2982 \div 6$.
20. $1 \times 1 \times 1 \times 1 \times 1 \times 1$.
21. $4 \times 7 \times 9 \times 0$.
22. 18×18.
23. 231×201.
24. £3 12*s.* + £1 4*s.* 6*d.* + 13*s.* 6½*d.*
25. £5 — £2 12*s.* 6*d.*
26. 2 hr. 35 min. 15 sec. + 46 min. 50 sec.
27. 3 lb. 4 oz. — 1 lb. 12 oz.
28. 3 yards 2 ft. 8 in. — 2 ft. 10 in.
29. 16*s.* 7½*d.* × 2.
30. £11 18*s.* ÷ 7.
31. £2 17*s.* 3*d.* × 5.
32. 1*s.* 3½*d.* × 15.
33. 12 lb. 3 oz. ÷ 5.
34. 7 weeks 4 days 8 hours × 3.
35. 2 ft. 3 in. × 8.
36. 5 yards 0 ft. 6 in. ÷ 6.
37. How much are 50 half-crowns?
38. Reduce £1 5*s.* 2*d.* to pence.
39. $\frac{1}{2} + \frac{1}{3}$.
40. $\frac{1}{3} - \frac{1}{4}$.
41. $\frac{2}{3} + \frac{3}{5}$.
42. $3\frac{5}{6} - 1$.
43. $6\frac{3}{5} - 4\frac{2}{5}$.
44. $1\frac{3}{4} - \frac{1}{2} + \frac{1}{4}$.
45. $2\frac{2}{3} + 1\frac{1}{2} - 2\frac{2}{3}$.
46. $\frac{3}{4} \times \frac{4}{5}$.
47. $\frac{2}{3} \div \frac{5}{6}$.
48. $\frac{3}{8} \times \frac{4}{5} \div \frac{3}{4}$.
49. $8 \times \frac{1}{2}$.
50. $8 \div \frac{1}{2}$.
51. £20 × ⅓.
52. $\cdot 5 + \cdot 75 + \cdot 25$.
53. $1 \cdot 2 + 3 \cdot 6 - 2 \cdot 4$.

54. 18·2 — 6·7 + 11·8.
55. 10 — 7·35.
56. 3·1 — 3·01.
57. ·1 × ·1.
58. ·6 × ·07.
59. 1 ÷ ·1.
60. 3 ÷ ·03.
61. 7·5 × 8 ÷ ·6.
62. ·011 × ·26.
63. 6 ÷ ·5.
64. $\sqrt{81}$.
65. $\sqrt{169}$.
66. $\sqrt{400}$.
67. Square root of 441.
68. $\sqrt{49} \times \sqrt{121}$.
69. 5 per cent. of 10*s*.
70. $3\frac{1}{2}$ per cent. of £300.
71. 1 per cent. of 100.
72. 100 per cent. of 1.
73. 100 per cent. of 100.
74. 200 per cent. of 100.
75. 3 lb. 4 oz. at 1*s*. a lb.
76. 4 yd. 1 ft. at 2*s*. 6*d*. a yd.
77. 32 oranges at 8 for 5*d*.
78. 100 oranges at 5 for 1*s*.
79. £3 12*s*. 6*d*. × $\frac{2}{3}$.
80. Reduce $\frac{21}{35}$ to lowest terms.
81. Find the value of £·375.
82. What decimal of £1 is 13*s*. 4*d*. ?
83. G.C.M. or H.C.F. of 32, 16, 56.
84. G.C.M. or H.C.F. of 12, 20, 28, 32.
85. L.C.M. of 1, 2, 3, 4, 5.
86. L.C.M. of 7, 8, and 112.
87. L.C.M. of 3, 5, 9, 15.
88. Average of 5, 17, 17, 19.
89. Average of $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$.
90. Average of ·5, ·01, 1·2.
91. 4 : 7 : : 6 : x.
92. 3 : x : : 5 : 10.
93. x . 8 : : 3 : 12.
94. 8 : 7 : : x : 21.
95. Area of plot $2\frac{1}{2}$ yd. sq.
96. Circum. of circle 4 ft. in diam. ($\pi = \frac{22}{7}$).
97. Area of circle 4 ft. diam.
98. Area of all the faces of a cube of 2 inch edge.
99. Volume of water in a tank 2 ft. long, 2 ft. wide, $1\frac{1}{2}$ ft. deep.
100. Find radius of circle with area of 154 sq. ft.

KEY TO MECHANICAL ARITHMETIC TEST

1. 13.
2. 23.
3. 28.
4. 15.
5. 16.
6. 21.
7. 2999.
8. 0.
9. 187.
10. 429.
11. 0.
12. 99.
13. 376.
14. 35.
15. 4109.
16. 1.
17. 5.
18. 0.
19. 497.
20. 1.
21. 0.
22. 324.
23. 46431.
24. £5 10*s*. 0½*d*.
25. £2 7*s*. 6*d*.
26. 3 hr. 22 min. 5 sec.
27. 1 lb. 8 oz.
28. 2 yd. 2 ft. 10 in.
29. £1 13*s*. 3*d*.
30. £1 14*s*. 0*d*.
31. £14 6*s*. 3*d*.
32. 19*s*. 4½*d*.
33. 2 lb. 7 oz.
34. 22 wks. 6 dys.
35. 6 yd.
36. 2 ft. 7 in.
37. £6 5*s*. 0*d*.
38. 302.
39. $\frac{5}{6}$.
40. $\frac{1}{12}$.
41. $1\frac{4}{15}$.
42. $2\frac{5}{6}$.
43. $2\frac{1}{5}$.
44. $1\frac{1}{2}$.
45. $1\frac{1}{2}$.
46. $\frac{3}{5}$.
47. $\frac{4}{5}$.
48. $\frac{2}{5}$.
49. 4.
50. 16.
51. £6 13*s*. 4*d*.
52. 1·5.
53. 2·4.
54. 23·3.
55. 2·65.
56. ·09.
57. ·01.
58. ·042.
59. 10.
60. 100.
61. 100.
62. ·00286.
63. 12.
64. 9.
65. 13.
66. 20.
67. 21.
68. 77.
69. 6*d*.
70. £10 10*s*. 0*d*.
71. 1.
72. 1.
73. 100.
74. 200.
75. 3*s*. 3*d*.
76. 10*s*. 10*d*.
77. 1*s*. 8*d*.
78. £1.
79. £2 8*s*. 4*d*.
80. $\frac{3}{5}$.
81. 7*s*. 6*d*.
82. ·6.
83. 8.
84. 4.
85. 60.
86. 112.
87. 45.
88. $14\frac{1}{2}$.
89. $\frac{1}{2}$.
90. ·57.
91. $10\frac{1}{2}$.
92. 6.
93. 2.
94. 24.
95. $6\frac{1}{4}$ sq. yd.
96. $12\frac{4}{7}$ ft.
97. $12\frac{4}{7}$ sq. ft.
98. 24 sq. in.
99. 6 c. ft.
100. 7 ft.

One Hour

ARITHMETICAL REASONING

Work the following sums in your head.

1. Jane is 18 years old. If Sarah were 5 years older she would be as old as Jane. How old is Sarah ?
2. What number is half-way between 12 and 16 ?
3. A man is taller than his wife by 3 inches; his wife is taller than his daughter by 5 inches. The daughter is 60 inches high. How high is the man ?
4. A boy was given 12 apples 3 of which were rotten and had to be thrown away. After eating 4 of the good ones how many were left ?
5. A boy has 3 miles to walk to get to school. He can cycle 4 times as fast as he can walk. How far has he to go when he cycles to school ?
6. Fred plays marbles starting with 15. He loses 8 and then wins 6. How many has he then ?
7. There are 20 beads on a string, of which 7 are red, 8 are blue, and the rest are yellow. How many are yellow ?
8. A plank 20 ft. long is laid on the top of one which is 14 ft. long so that it is 2 ft. over at one end. How much is it over at the other ?
9. John has 30 marbles and James 15. How many more will John have than James after he has given James 6 of his own ?
10. Among how many boys can I share half a crown so that each gets 3*d.* ?
11. I make as many separate triangles as I can with 29 whole matches; how many matches are not used ?
12. If I write the word "giraffe" 8 times how many letters shall I make ?
13. If in one scale of a balance there are 15 lb. of rice and in the other 7 lb. of apples, how much rice must I take from one pan and place in the other to make them balance ?
14. Two similar taps fill a bath in 20 minutes; how long will one take ?
15. Mary is twice as old as Jane, Jane is twice as old as Annie, Annie is as old as Ruth, Ruth is 3; how old is Mary ?

16. There are 40 nuts on a plate; how many will be left after 5 people have each eaten 7 ?
17. Five separate equilateral triangles of equal size are made from 60 in. of wire; how long is each side ? (An equilateral triangle is one which has three equal sides.)
18. A man takes 20 minutes to walk from his house to the station. His son also takes 20 minutes. How long will it take them if they both walk together ?
19. There are 35 boys in the first class and 40 in the second. For the arithmetic lesson 5 boys go down from the first class and 7 go up from the second to the first. How many boys are in the first class for the arithmetic lesson ?
20. Find the number a quarter of which is 5.
21. The bottom of a hill is 200 ft. above the level of the sea and the top 400 ft. above the level of the sea. How high above the sea is a house half-way up the hill ?
22. Shakespeare died in 1616 at the age of 52. When was he born ?
23. What is the length of a stick which I can cut up into 8 pieces each 6 in. long, and have 4 in. left over ?
34. How much must be added to 4*s.* 6*d.* to make 10*s.* ?
25. Find the cost of 9*d.* worth of eggs at 7 for 6*d.*
26. If I pay 6*s.* for 3 lb. of butter, what shall I have to pay for 2 lb. ?
27. How many numbers between 19 and 30 are exactly divisible by 4 ?
28. If it takes 3 minutes to boil an egg, how long will it take to boil 10 eggs together ?
29. How many twos must be multiplied together to make 32 ?
30. A man bought on January 4th 10 lb. of chestnuts for 3*s.*, and sold them in the street on January 6th to 30 customers for 5*s.* What profit did he make ?
31. A boy measuring a piece of string with a foot-rule found it to be 6 ft. long; but somebody had cut an inch off the rule. What was the real length of the string ?
32. A paper boy buys a dozen newspapers at 1¼*d.* each and sells them at 1½*d.* each. What profit does he make on the whole ?

33. If I buy two books at 5*d.* each and three tops at 3*d.* each what change shall I have left from a two-shilling piece ?
34. What is the smallest sum of money that can be paid in either two-shilling pieces or half-crowns ?
35. A boy was given half a crown. He gave half to his brothers and with the rest bought threepenny bars of chocolate. How many bars did he buy ?
36. A butcher in selling 3¼ lb. of meat gave short weight to the extent of 5 oz. What did the meat really weigh ?
37. A runs a hundred yards race with B, giving him 5 yards start. A beats B by 6 yards. How many yards has B run when A reaches the winning post ?
38. If a train runs 12 miles in 10 minutes, how long will it take to run a mile ?
39. A man has 5 children. It costs him £2 10*s.* a week to feed them. What will be the expense for a month ?
40. If 1½ lb. of cheese costs 3*s.*, what will 1 lb. cost ?
41. A man of 35 is 7 times as old as his son ; how many times as old as his son will he be 25 years hence ?
42. Two-thirds of a class consists of 24 children. How many are there in the class ?
43. A half and a quarter of a man's money added together make 9*s.* How much has he ?
44. If half a cake costs 2½*d.* what will 5 cakes cost ?
45. A boy spent a quarter of his money on sweets and twice as much on fruit. Half of what he had left was 1½*d.* How much had he at first ?
46. Among how many boys may 35 apples be divided so that each gets 2½ apples ?
47. How many days are from noon on January 28 to noon on February 2 ?
48. If five-eighths of my money is 10*d.*, how much money have I ?
49. After spending half of my money and then half the remainder I had 2*d.* left. How much had I at first ?
50. I entered a shop at 5 p.m. and stayed till 5.30 p.m. I bought two ties at 2*s.* 6*d.* each and two handkerchiefs at 3*s.* 6*d.* each. What change did I get from £1 ?
51. A boy sold a knife for 1*s.* 6*d.*, gaining 3*d.* on what he gave for it. What fraction of the cost price did he gain ?

52. What is the least number that must be added to 53 to make it exactly divisible by 7 ?
53. How many eggs at 3 for 5*d.* can I buy for 2*s.* 6*d.* ?
54. How many oranges at 4 for 3*d.* can I buy for 2*s.* ?
55. If it takes 3 men to paint the inside of a house in two days, how many men would be needed to do it in half a day ?
56. Two men start rowing down a river from the same pier at the same time. One rows at the rate of 4 miles an hour, and the other at 3¼ miles an hour. How far apart are they after 3 hours ?
57. My watch gains 4 minutes every day. If it is set right at noon on Monday what time will it show on the following Wednesday when the right time is 6 p.m. ?
58. If half an acre of land costs £15, what will a third of an acre cost at 100 per cent. of the same price ?
59. In the 3 class-rooms of a school there are 120 children. Half of them are in the first room, one-third in the second room, and the rest in the third room. How many are there in the third room ?
60. A bag contained nuts. Dick was given half of them and Sam a quarter. It was found that Dick had 9 nuts more than Sam. How many nuts were there in the bag at first ?
61. What fraction lies midway between ¼ and ½ ?
62. What two whole numbers multiplied together will make 7 ?
63. Find the distance round a square whose side is 8 in.
64. Maud and Gladys together earn 10*s.* by making button-holes. Maud sews 3 while Gladys sews 2. What share should Maud get ?
65. What is the greatest number that will divide 62 and 74 and leave a remainder of 2 in each case ?
66. A 5-story house has 4 equal flights of stairs with a total of 52 stairs. How many stairs must I go up to get to the fourth story ?
67. What is the profit on 120 articles bought at 8*d.* a dozen and sold at 1*s.* a dozen ?
68. A bookseller bought some books for £1 and sold them for £1 2*s.* 6*d.* gaining 6*d.* on each book. How many did he buy ?

69. A man 6 ft. high stands in the sunlight near a telegraph pole 30 ft. high and casts a shadow of 9 ft. How long is the shadow of the telegraph pole ?

70. If 1 man eats 1 apple in 1 day, in how many days will 10 men eat 10 apples ?

71. A man can dig his garden in 2 days, and his son can do it in 4 days. How long will it take them if they work together ?

72. How much is three times the third of three and a third ?

73. The average of 5 numbers is 4. Four of the numbers are 2, 1, 5, and 7. What is the fifth number ?

74. There are 2 numbers one of which is larger than the other by 2. When multiplied together they make 143. What is the smaller number ?

75. A man works a day, then rests a day, then works a day, then rests a day, and so on. For each day he works he earns 15*s*. How much will he earn from Monday morning to Friday night ?

76. How many times as fast as the hour hand does the minute hand of a clock move ?

77. The area of a square is 144 sq. in. What is the distance round it ? (In inches.)

78. What is the distance round an oblong table 5 ft. long and 4 ft. broad ?

79. A rectangle is twice as long as it is broad. Its area is 200 sq. ft. What is its breadth ?

80. If I share 6*d*. among two boys, so that one gets 2*d*. more than the other, how much does the less fortunate boy get ?

81. If I share a shilling among two boys so that one gets a penny more than the other, how much does the more fortunate boy get ?

82. A flower-bed 6 ft. square is surrounded by a path a foot wide. Find the area of the path in square feet.

83. Seven posts 3 ft. apart are fixed in a row. How far is the first post from the last ?

84. A boy said: "In 10 years' time I shall be twice as old." How old was he ?

85. If telegraph poles stand 50 yards apart in a straight row, what is the distance from the first to the eighth ?

86. John and Henry start walking to meet one another from places 10½ miles apart. If John walks at the rate of 3 miles an hour and Henry at 4 miles an hour, how long will it be before they meet ?

87. A man walking at the rate of 4 miles an hour pursues another man who has had an hour's start and walks at the rate of 3 miles an hour. How long will it take the pursuer to catch the other man ?

88. A rectangular plot of grass 6 ft. long and 5 ft. broad is surrounded by a path 2 ft. wide. What is the area of the path in square feet ?

89. Five per cent. of A's income is the same as 15 per cent. of B's. A's income is £300 a year. What is B's ?

90. A street 30 yards long is planted on each side with trees 6 yards apart. How many trees are there ?

91. If a man's salary is increased by 10 per cent. and then reduced by 10 per cent., state whether he loses or gains and by what percentage.

92. If a man's salary is reduced by 10 per cent. and then increased by 10 per cent., state whether he loses or gains and by what percentage.

93. A box and its key cost 1*s.* 2*d.* The box costs 1*s.* more than the key. What does the key cost ?

94. The sum of 2 numbers is 24 and their difference is 10. What is the larger of the 2 numbers ?

95. A man rows with the stream at the rate of 3 miles an hour and against it at the rate of 1 mile an hour. What is the rate of the stream ?

96. A brick weighs 7 lb. and half its own weight. What is the weight of the brick ?

97. How many times can one-third be taken away from 12 ?

98. A man walks 5 yards to the north, 5 yards to the east, 5 yards to the south, and then 5 yards to the west. How far is he then from the starting point ?

99. If in secretly sending numbers to a friend I agree to write 7 when I mean 3 and 11 when I mean 7, what should I write when I mean 10 ?

100. The first even number is 2, the second is 4, and so on. What is the hundredth even number ?

KEY TO ARITHMETICAL REASONING TEST

1. 13.
2. 14.
3. 68 in.
4. 5.
5. 3 miles.
6. 13.
7. 5.
8. 4 ft.
9. 3.
10. 10.
11. 2.
12. 56.
13. 4 lb.
14. 40 min.
15. 12.
16. 5.
17. 4 in.
18. 20 min.
19. 37.
20. 20.
21. 300 ft.
22. 1564.
23. 4 ft. 4 in.
24. 5*s.* 6*d.*
25. 9*d.*
26. 4*s.*
27. 3.
28. 3 min.
29. 5.
30. 2*s.*
31. 5 ft. 6 in.
32. 3*d.*
33. 5*d.*
34. 10*s.*
35. 5.
36. 2 lb. 15 oz.
37. 89 yd.
38. 50 sec.
39. £10.
40. 2*s.*
41. Twice.
42. 36.
43. 12*s.*
44. 2*s.* 1*d.*
45. 1*s.*
46. 14.
47. 5.
48. 1*s.* 4*d.*
49. 8*d.*
50. 8*s.*
51. $\frac{1}{5}$.
52. 3.
53. 18.
54. 32.
55. 12.
56. $2\frac{1}{4}$.
57. 6–9 p.m.
58. £10.
59. 20.
60. 36.
61. $\frac{3}{8}$.
62. 7 and 1.
63. 32 in.
64. 6*s.*
65. 12.
66. 39.
67. 3*s.* 4*d.*
68. 5.
69. 45 ft.
70. 1.
71. $1\frac{1}{2}$ days.
72. $3\frac{1}{3}$.
73. 5.
74. 11.
75. 45*s.*
76. 12 times.
77. 48 in.
78. 18 ft.
79. 10 ft.
80. 2*d.*
81. $6\frac{1}{2}$*d.*
82. 28 sq. ft.
83. 18 ft.
84. 10 yrs.
85. 350 yd.
86. $1\frac{1}{2}$ hr.
87. 3 hr.
88. 60 sq. ft.
89. £100.
90. 12.
91. Loses 1 %.
92. Loses 1 %.
93. 1*d.*
94. 17.
95. 1 m.p.h.
96. 14 lb.
97. 36 times.
98. 0.
99. 14.
100. 200

Two Hours
Scrap paper allowed

ALGEBRA

1. How many shillings are there in x pence ?
2. If you are y years old now how old will you be in 2 years' time ?
3. The side of a square is p inches long. What is the distance round it ?
4. If a train travels at the rate of n miles an hour, how far will it go in t hours ?
5. A room is m feet long and n feet broad. Find its area in square yards.
6. If a pencil costs a pence, how many can be bought for b pence ?
7. What must be subtracted from x to give y ?
8. What must be divided by p to give q ?
9. If $x = 10$ what is the value of $3x^2 + 4x + 6$?
10. If $x = 10$ what is the value of $5x^3 + 8x + 2$?
11. In the series 1, 2, 3, 4, etc., what number comes before n ?

If $a = 2$, $b = 1$, $c = 3$, $d = -2$, and $e = 0$, find the value of the following 8 expressions :

12. $a(c - b)$.
13. $4e^3 + 3e^2 + 1$.
14. $3a(b - c)$.
15. $a^2 - d^2$.
16. $\sqrt{3c(a + b)^2}$.
17. $a - 3d + e$.
18. $ab - (e - d)$.
19. $\dfrac{b^2 + c^2}{ac - ad}$.

Write the following 9 expressions in their simplest form (do not factorise) :

20. $4ab - 5ab$.
21. $3x \times 5x$.
22. $4x^3 \times 2x^2$.
23. $2a + b + c$.
24. $4x^3 \div 2x^2$.
25. $4x^3 - 2x^2$.
26. $3x \div 5x$.
27. $a^m \times a^n$.
28. $a^m \div a^n$.

Expand the following 9 expressions :

29. $3x(x - y)$.
30. $-2(2x - 1)$.
31. $(x + 1)(2x + 3)$.
32. $(3m + 2n)(3m - 2n)$.
33. $(2a - 3)(3a - 2)$.
34. $(4y + 1)(2y - 5)$
35. $(2a + 3b)^2$.
36. $(x - 2y)^2$.
37. $(3a - 2a)^2$.

Factorise the following 10 expressions :

38. $3 + 6x$.
39. $a^2b - 3a^3$.
40. $a^2 + 2a + 1$.
41. $4x^2 - y^2$.
42. $m^{10} - n^8$.
43. $a^2 + 5ab + 6b^2$.
44. $3x^2 - 10x + 3$.
45. $2a^2 + 3ab - 2b^2$.
46. $4x^2y^2 + 4xy - 3$.
47. $7x + 2y)^2 - 2(x + 2y)$

What values of x will make the following 10 statements true?

48. $4x - 3 = 5$.
49. $x^2 = 36$.
50. $\frac{1}{x} = 3$.
51. $(x - 1)^2 = 0$.
52. $\sqrt[3]{27} = 3$.
53. $x^2 + x = 0$.
54. $2^x = 16$.
55. $(x - a)(x - b) = 0$.
56. $ax + b = c$.
57. $3(x - a) = 3x - 3a$.

58. Solve the equation: $\frac{4x - 1}{5} = 3$.

59. Solve the equation: $4 + \frac{3}{x} = \frac{6}{x} - 1$.

60. Solve: $x + y = 11$.
$x - y = 5$.

61. Solve: $3x - y = 1$.
$2y - x = 8$.

62. Solve: $3x - 5y = 4$.
$x - 4y = -1$.

63. Half the sum of two numbers is 7 and half their difference is 2. Find the numbers.

64. A man's age is double that of his son; 15 years ago it was three times that of his son. How old is the man?

Without altering the first two terms state what must be added to or subtracted from the following five expressions to make them perfect squares:

65. $x^2 + 4x$.
66. $a^2 - 3a$.
67. $p^2 + 6p + 2$.
68. $r^2 + 8r + 20$.
69. $yt - y^2$.

70. Which is the larger, (*a*) the square of the sum of two numbers or (*b*) the sum of their squares?

If the former put down (*a*); if the latter put down (*b*); if they are equal put down =.

71. The area of a rectangle is 21 sq. in., and its perimeter is 20 in. What are its sides?

72. Find two consecutive numbers the difference of whose squares is 11.

73. The sum of a number and its square is 42. What is the number?

74. Simplify: $\left\{\frac{4a^2x - 3by^2}{4a^2x - 3by^2}\right\}^8$

75. Solve. $x^2 - 5x + 6 = 0$.

76. Solve: $3x^2 - 6x = 9$.

77. Simplify: $2 - \frac{2(x-1)}{3}$.

78. Simplify: $\frac{x}{y} + \frac{y}{x}$.

79. Simplify: $\frac{p}{p-q} - \frac{q}{p+q}$

80. Simplify: $\frac{1}{a-b} - \frac{1}{b-a}$.

81. Simplify: $\frac{a^2}{a^2-b^2} \times \frac{a-b}{a}$.

82. Simplify: $\frac{a^2-b^2}{c^2-d^2} \div \frac{b-a}{c+d}$.

83. Solve: $\frac{x}{a} + \frac{x}{b} = a + b$.

84. Solve: $a(x+b) - b(x+a) = a^2 - 2ab + b^2$.

85. Solve: $(a+x)(b+x) = 0$.

86 Solve: $\left\{x - \frac{a+b}{b}\right\}\left\{x + \frac{c+d}{d}\right\} = 0$.

87. Solve: $3(2x-3) + 5x(2x-3) = 0$.

88. Form the equation whose roots are 3 and 4.

89. Form the equation whose roots are a and $-b$.

90. Simplify: $\left\{\frac{a+b-(a+b)}{x+y}\right\}^6$

91. Express this without the brackets: $(-2xy^2)^3$.

92. Simplify: $\sqrt{3}\sqrt{12}$.

93. Simplify: $\sqrt{12} + \sqrt{27}$.

94. Simplify: $\sqrt{a^5b^4}$.

95. If α and β are the roots of the equation $px^2 - qx + r$ what is the value of $\alpha + \beta$?

96. What is the value of the nth term in the series 2, 4, 6, 8, 10, etc?

97. What is the value of the nth term in the series 1, 3, 5, 7, 9, etc.?

98. What is the value of the $(n-1)$th term in the series $\frac{1}{2}$, 1, $1\frac{1}{2}$, 2, $2\frac{1}{2}$, etc.?

99. A stone dropped from a balloon falls 16 feet in the first second, 4×16 ft. in two seconds, 9×16 ft. in three seconds, etc. How far will it fall in n seconds?

100. Find the value of $a^0 + b^0 - c^0$.

KEY TO ALGEBRA TEST

1. $\frac{x}{12}$
2. $y + 2$.
3. $4p$.
4. nt miles.
5. $\frac{mn}{9}$
6. $\frac{b}{a}$
7. $x - y$.
8. pq.
9. 346.
10. 5082.
11. $n - 1$.
12. 4.
13. 1.
14. -12.
15. 0.
16. 9.
17. 8.
18. 0.
19. 1.
20. $-ab$.
21. $15x^2$.
22. $8x^5$.
23. $2a + b + c$.
24. $2x$.
25. $4x^3 - 2x^2$.
26. $\frac{3}{5}$.
27. a^{m+n}
28. a^{m-n}
29. $3x^2 - 3xy$.
30. $-4x + 2$.
31. $2x^2 + 5x + 3$.
32. $9m^2 - 4n^2$.
33. $6a^2 - 13a + 6$.
34. $8y^2 - 18y - 5$.
35. $4a^2 + 12ab + 9b^2$.
36. $x^2 - 4xy + 4y^2$.
37. a^2.
38. $3(1 + 2x)$.
39. $a^2(b - 3a)$.
40. $(a + 1)^2$.
41. $(2x + y)(2x - y)$.
42. $(m^5 + n^4)(m^5 - n^4)$.
43. $(a + 3b)(a + 2b)$.
44. $(3x - 1)(x - 3)$.
45. $(2a - b)(a + 2b)$.
46. $(2xy - 1)(2xy + 3)$.
47. $(x + 2y - 2)(x + 2y)$.
48. 2.
49. ± 6.
50. $\frac{1}{3}$.
51. 1.
52. 3.
53. 0 or -1.
54. 4.
55. a, b.
56. $\frac{c - b}{a}$
57. x = any value.
58. 4.
59. $\frac{3}{5}$.
60. 8, 3.
61. 2, 5.
62. 3, 1.
63. 9, 5.
64. 60 years.
65. Add 4.
66. Add $\frac{9}{4}$.
67. Add 7.
68. Add -4 or subtract 4.
69. Add $\frac{1}{4}t^2$.
70. (a).
71. 3″ and 7″.
72. 5 and 6.
73. 6.
74. 1.
75. $x = 2$ or 3.
76. $(3x + 3)(x - 3) = 0$,
i.e. $x = -1$ or 3.
77. $8 - 2x$.
78. $\frac{3}{xy}$.
79. $\frac{p^2 + q^2}{p^2 - q^2}$.
80. $\frac{2}{a - b}$.
81. $\frac{a}{a + b}$.
82. $-\frac{a + b}{c - d}$.
83. ab.
84. $a - b$.
85. $-a, -b$.
86. $\frac{a + b}{b}, -\frac{c + d}{d}$.
87. $\frac{3}{4}, -\frac{3}{5}$.
88. $x^2 - 7x + 12 = 0$.
89. $(x - a)(x + b) = 0$.
90. 0.
91. $-8x^3y^6$.
92. 6.
93. $5\sqrt{3}$.
94. $a^{\frac{2}{3}}b^2$.
95. $\frac{q}{p}$.
96. $2n$.
97. $2n - 1$.
98. $\frac{n - 1}{2}$
99. $16n^2$.
100. 1.

CHAPTER XVIII

GEOGRAPHY

WHAT are the fundamentals of Geography? The question would be answered one way thirty years ago and another way to-day. And even to-day it is impossible to give an answer which would secure universal assent. Opinions shift and change. Time was when Geography was mainly concerned with bare facts: it dealt with boundaries and areas and populations, with the names of places, the length of rivers, and the height of mountains. And it described things according to formula, and with strict impartiality. All the counties of England were equal in the eyes of the geographer; each had to be learnt with its county town, together with the river on which it stood. And Oakham was a troublesome little town that spoiled the completeness of the list. And there was nothing to show that one fact was more momentous than another. The fact that Bodmin is on the Camel (if it really is on the Camel) was just as important as the fact that London is on the Thames or that Russia is on the rocks.

This was the old Geography—the Geography of facts. A great change, however, took place about a quarter of a century ago. Geography passed from the factual stage to the rational stage. Everything had to be explained. The reason why was the

dominant quest. Facts were presented, not in isolation, but as dependent on one another, as bearing to one another the relation of cause and effect. A natural region rather than a political area became the unit of study, and attempts were made to trace the origin of the dwellers in those parts—of all that they are and of all that they do—to the geology of the region, or its climate, or the ocean currents that passed its shores. The primal cause was, in fact, found in the heavens above or in the earth beneath or in the waters under the earth. And the proper nouns with which the old Geography was so painfully congested began to disappear. In some cases they disappeared altogether. We arrived at a Geography without names. I well remember seeing many years ago the work of an ardent disciple of the new school. He prided himself on having taken his pupils through a year's course in the Geography of England and Wales without mentioning a single geographical name. He had begun farther back than the Earth: he had begun with the sun and the solar system and had worked his way gradually to this our little lump of rock "set in the silver sea." And the boys could do wonders with contours and sections and wind-charts and isobars and isotherms and other fearsome things. But if they had ever heard of Sheffield, or Exeter, or Bangor, it was purely by accident. The teacher never intended to let it out.

It is now recognised that the rationalisation of Geography was overdone. We cannot reason *in vacuo*: we must have facts to start from. We can show a logical connection between some facts, but we cannot show a logical connection between all

facts. There are such things as data in Geography —things given, not things derived. We have to swallow the fact that Birmingham is a big town in the Midlands, and argue about it afterwards if we like. In fine, facts must be taught as well as reasons; names must be known as well as places. There is, after all, no great merit in not knowing the name of a place.

We must distinguish, however, between fact and fact. The new examiner acts upon the principle that fundamental things should be tested more frequently and more searchingly than accessory things—that difficulty as a basis of examination should give place to importance. But what are the important things? What is the irreducible minimum of Geography that a lad of fourteen should know when he leaves school? What are the things that a man of culture cannot forget—that his intercourse with the world of business or pleasure, of politics or letters, will not allow him to forget? What is necessary to his clear understanding of the geographical references in newspapers and books? If we can discover these things we can discover the essentials of which we are in search. But can we discover them? We can when we can foretell what will appear in to-morrow's paper. Who in the year 1913 could have foreseen that Serajevo was ever to assume the tremendous significance it did in 1914. Ten years ago Ypres was known only to the few who could pronounce it; to-day it is known to the multitude and pronounced in a myriad ways.

Still, it may reasonably be urged that these are historical, and not geographical, vicissitudes. Geo-

graphically these places are of no more significance than before. Indeed, Serajevo has already returned to its former obscurity as a sleepy little town in the Near East; while Ypres is slowly passing into history, and will in course of time share the geographical fate of Runnymede, Blenheim, and Waterloo. All this may readily be conceded; but the examples show at least one thing: they show that we cannot rely upon the newspapers for our geographical values.

Must we then fall back upon the textbooks in current use? There are obvious objections. If the textbooks of the last generation are discredited by our own generation, what certainty have we that the textbooks of our generation will not be discredited by the next generation? Besides, the textbook criterion has proved worthless in other fields—particularly in spelling. The new examiner has discarded the spelling-book and followed the line indicated by Ayres, who inquired into the frequency with which words were used in every-day life, and based a spelling scale upon the results of that inquiry. He thus arrived at a scale of examination values which could not be challenged. But this is a method which cannot be pursued in the testing of Geography.

The Geography examination that follows does not by any means meet the demands I have just described. It does not pretend to be anything more than an example of the new technique of examining as applied to the geography of England and Wales. The subject-matter was selected from the text-books in use in the schools. I compared the contents of the best of these books and noted the

common factors. It is upon these common factors that my questions are based. Defective as this method is, I know of no better. It at least has the merit of bringing the examination closely into line with the teaching. I have chosen a part of the geographical field which cannot be quite overlooked, and yet is peculiarly liable to be neglected—in the elementary school at any rate. The ordinary boy leaves school knowing more about the geography of India than about the geography of his own land. The tradition is for the formal course to begin with England and Wales [1] in Standard III and to pass on to other countries in the higher standards. Thus the young scholar deals with his own country when his mind is least able to grasp the significance of geographical facts and of geographical reasoning. And after the age of ten he studies it no more. Whether geography should begin at home or not is a question which we need not here dispute. It is, no doubt, as good a place to begin at as anywhere else. But it is scarcely wise to take the pupils abroad and never bring them home again. If we begin with England we must necessarily treat it in a very simple and sketchy way. And the sketch needs filling in during the last year of school life. It is for pupils who follow such a course as this that my test was devised. I have not found many schools where the geography of England and Wales is revised in the top class. Where it is, and my examination has been set, the average mark obtained is about 50 out of a total of 100. The semi-inter-

[1] My colleague, Mr. Leonard Brooks, points out that the British Isles constitute a better unit for teaching purposes than England and Wales.

quartile range or quartile deviation is about 10; that is to say, half the pupils get marks which range between 40 and 60. The other half get marks which are either higher than 60 or lower than 40.

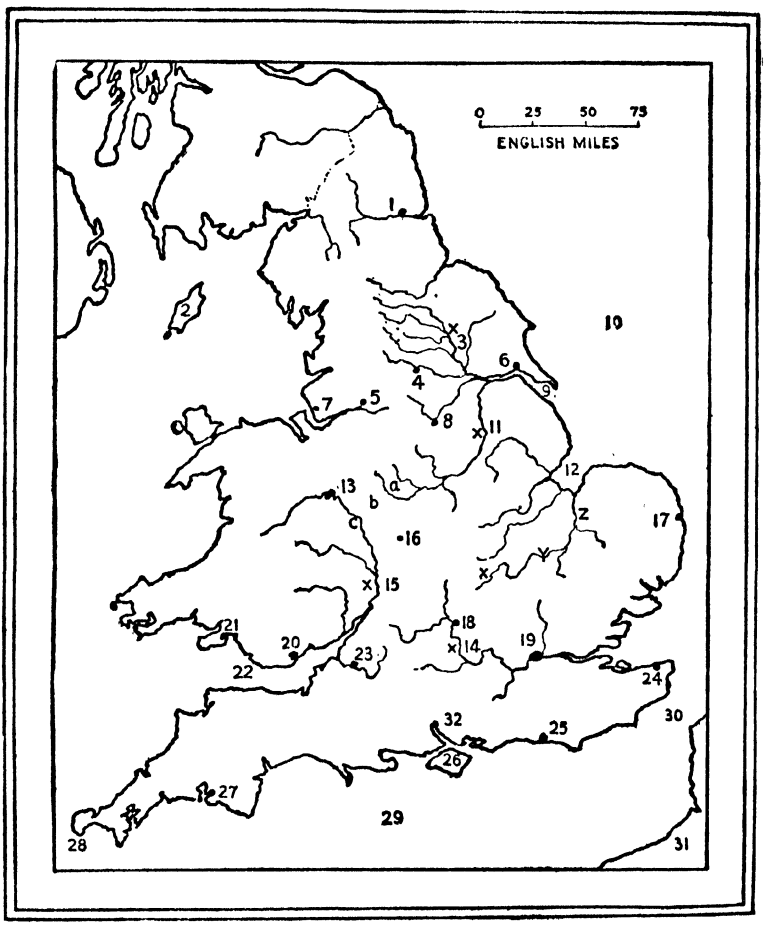
When the geography of England and Wales is specially studied in central schools the median mark varies with the year. It is about 50 in the first year and 70 in the fourth. The quartile deviation is less than in the ordinary school.

Special care should be taken in marking the paper to keep the score for the true-false test within its own province. It must stand by itself and not be allowed to carry discredit to the other part of the paper. Since the score for the true-false test is R — W, if there are more wrong answers than right there will be a negative score which will reduce the total obtained on the other items. Negative marks are, however, to be ignored: no pupil is credited with less than zero. Hence, we cannot arrive at the correct score by counting the number of items right in the whole paper, and then deducting the number of items wrong in the true-false section. We may be deducting too many. It is consequently safest to record the score on each paper in this way:

	Marks
Items 1 to 55	
Items 56 to 100 (R — W) . . .	
Total . .	

Answers to questions (33), (34), and (35) are considered right if the distances given are within 10 miles on either side of the true distances.

ENGLAND AND WALES



Forty-five minutes

THE GEOGRAPHY OF ENGLAND AND WALES

(Ruler or strip of paper allowed for questions 33, 34, and 35.)

In the accompanying map numbers are put instead of the names of places. If the number is quite close to a dot a town is meant, if close to a cross, a river. You are asked to put down on your answer paper the number that stands for the places named in the first thirty questions. If the first name were "Margate" you should give the number (1) 24. If the second name were "English Channel" you should give the number (2) 29. Now give the number of these places:

(1) London.
(2) Hull.
(3) River Severn.
(4) Newcastle.
(5) Straits of Dover.
(6) Bristol.
(7) Isle of Man.
(8) River Thames.
(9) Land's End.
(10) Shrewsbury.
(11) Yarmouth.
(12) Liverpool.
(13) North Sea.
(14) Plymouth.
(15) Leeds.
(16) The Wash.
(17) Swansea.
(18) Isle of Wight.
(19) Oxford.
(20) The Humber.
(21) Cardiff.
(22) France.
(23) River Ouse.
(24) Manchester.
(25) Bristol Channel.
(26) Sheffield.
(27) Southampton.
(28) River Trent.
(29) Brighton.
(30) Birmingham.

(31) Of the three places marked *x*, *y*, and *z* on the map, which stands the highest above the level of the sea?

(32) Of the three places marked *a*, *b*, and *c*, which stands the highest above the level of the sea?

(33) Find the distance as the crow flies between the places marked 19 and 24.

(34) Find the distance as the crow flies between the places marked 19 and 1.

(35) Find the distance as the crow flies between the places marked 1 and 27.

In the following statements wherever a number appears there is a missing word or two words, if they form a single name. You have to supply the missing words.

Sheffield is famous for its (36).

The chief industry of the South Lancashire coal-field is the manufacture of (37) goods, largely because its ports face those of (38), which originally supplied all the raw (39).

The line joining the mouth of the Exe with the mouth of the Tees divides England and Wales into two parts, a (40) region to the north-west and a (41) region to the south-east.

The backbone of England is the (42) which runs in the direction (43) to (44).

Leeds manufactures (45) goods because it is near good grazing ground, is situated in the (46) coal-field and is on a tributary of the river (47), the water of which is used for cleansing and dyeing purposes.

Devonshire is noted for its (48) because it has fine (49) orchards.

The (50) canal, navigable by ocean steamers, connects Liverpool with Manchester.

Much warmth and rain are brought to England by the (51) wind.

England is near many countries with which she can trade because she is in the centre of the (52) hemisphere.

One of the two older universities is at (53) on the Thames; the other is at (54).

The highest mountain in England and Wales is called (55).

Some of the following statements are true and some are false. Read each carefully. If you think it is true put "Yes" on your answer paper; if you think it is false put "No." If you don't know put a dash ——. I will now give three statements and show you how to deal with them.

(1) Scotland is to the north of England. (1) Yes.

(2) Brighton is the name of a mountain. (2) No.

(3) There are more sheep than cattle in Middlesex. (3) ——.

Deal in this way with the following forty-five statements. (No statement must be regarded as true unless every bit of it is true.)

(56) Our imports are mainly food and raw materials for manufacture.

(57) The Potteries on the North Staffordshire coal-field make earthenware goods, partly from local clay and partly from Cornish clay.

(58) The river Tyne flows in a valley between the Cambrian Mountains and the Pennines.
(59) The Furness district in Lancashire produces fine iron ore.
(60) Because moisture is needed to prevent cotton yarn from breaking cotton manufacture flourishes east of the Pennines.
(61) Southampton is the second largest port in England.
(62) York is an old Roman town.
(63) Boots are made at Stafford, partly because it is in the midst of a grazing district.
(64) Mackintoshes are used more in Kent than in Cumberland.
(65) A white collar keeps clean longer in Devonshire than in Staffordshire.
(66) The shortest route to the Continent is across the North Sea from Harwich to Ostend.
(67) Garden vegetables are grown in large quantities around the big towns, especially London.
(68) The two largest forests in England are the New Forest and the Forest of Dean.
(69) The most important mineral found in England is coal.
(70) The best slates come from Cornwall.
(71) Sheep that graze on the drier moors and downs produce the better wool, and those that graze on the damp mountains produce the better mutton. Therefore Welsh wool is better than Welsh mutton.
(72) As a port, London has the drawback of having no coal or iron in the neighbourhood.
(73) The most ancient parts of London lie south of the Thames.
(74) Wheat, fruit, and hay ripen better in the dry east than in the moist west.
(75) The South Wales coal-field supplies brittle coal for the household.
(76) Generally speaking the population is most scanty on the coal-fields.
(77) The export of manufactured goods is the main source of England's prosperity.
(78) Linen goods of various kinds are made in the Black Country towns near Birmingham.
(79) If the surface of the seas round the British Isles were to fall 600 feet England would be joined to the Continent and Ireland would be joined to England.

(80) The winds that most frequently blow over England come from the north and north-west.
(81) We live in an island with a coast-line short in comparison with the area of the island.
(82) The heaviest rainfall is in the west.
(83) The oldest rocks are in the east.
(84) Coal-fields are exposed on both sides of the Pennines.
(85) The longest rivers flow mainly towards the west.
(86) In the Wash the tide rises and falls more than in the Bristol Channel.
(87) The lakes are found among the plains.
(88) As the chief watershed is near the west coast the rivers flowing towards the west are shorter and more rapid than those flowing towards the east.
(89) As the source of the Thames is only about 300 feet above the sea, and the river is very long, it is excellent for navigation.
(90) The west of England has greater extremes of heat and cold than the east.
(91) The interior has more sunshine than the sea-coast.
(92) No part of England is far from the sea; therefore it has a wide range of temperature.
(93) All the great railway systems radiate from Birmingham.
(94) Most of the water that Londoners drink comes from the Thames and its tributary the Lea.
(95) In passing through Devonshire and Cornwall by train we notice that vegetation is more luxuriant in Devonshire than in Cornwall. This is due to a difference in the soils which again is due to a difference in the rocks underneath.
(96) The fumes of copper-smelting kill vegetation; hence we find near Swansea large expanses of country which are brown and bare.
(97) London stands on a thick layer of clay.
(98) In the south-east of England most of the hills are made of granite.
(99) The shipbuilding industry flourishes at the mouth of the Tyne because coal and iron are found in the neighbourhood, and timber can easily be brought from Scandinavia across the North Sea.
(100) London began to grow round Blackfriars Bridge.

KEY TO GEOGRAPHY TEST

1. 19.
2. 6.
3. 15.
4. 1.
5. 30.
6. 23.
7. 2.
8. 14.
9. 28.
10. 13.
11. 17.
12. 7.
13. 10.
14. 27.
15. 4.
16. 12.
17. 21.
18. 26.
19. 18.
20. 9.
21. 20.
22. 31.
23. 3.
24. 5.
25. 22.
26. 8.
27. 32.
28. 11.
29. 25.
30. 16.
31. *x*.
32. *b*.
33. 60 (50–70).
34. 228 (218–230).
35. 295 (285–305).
36. cutlery.
37. cotton.
38. America.
39. cotton, material.
40. Mountainous, highland, mining.
41. flat, lowland, agricultural.
42. Pennine Range
43. North.
44. South.
45. woollen.
46. Yorkshire.
47. Ouse.
48. cider.
49. apple.
50. Manchester, ship.
51. South-west.
52. land.
53. Oxford.
54. Cambridge.
55. Snowdon.
56. Yes.
57. Yes.
58. No.
59. Yes.
60. No.
61. No.
62. Yes.
63. Yes.
64. No.
65. Yes.
66. No.
67. Yes.
68. Yes.
69. Yes.
70. No.
71. No.
72. Yes.
73. No.
74. Yes.
75. No.
76. No.
77. Yes.
78. No.
79. Yes.
80. No.
81. No.
82. Yes.
83. No.
84. Yes.
85. No.
86. No.
87. No.
88. Yes.
89. Yes.
90. No.
91. No.
92. No.
93. No.
94. Yes.
95. Yes.
96 Yes.
97. Yes.
98. No.
99. Yes.
100. No.

CHAPTER XIX

HISTORY

THERE is danger as well as difficulty in issuing a standard test in a subject like History where the body of material is huge in bulk and vague in outline. The danger is that the test should become a syllabus—that what is intended to measure should be used to limit and to stereotype. Hence, a standard test should always deal with minima. It should be concerned with seeing if the foundations are sound. And by the foundations I mean those elemental facts and those broad principles of interpretation about which there is common agreement. Nobody doubts that the invention of the steam-engine made a great difference to England; nobody doubts that the defeat of Napoleon at Waterloo made a great difference to Europe; and nobody doubts that these are among the facts that every child in our school should know before he goes out into the world. If a standard test contains no other things than these and such as these the danger disappears. For it does not in the least matter if such a test is used as a syllabus; for at the most it can be nothing but the mere skeleton of a syllabus: the flesh and blood will still have to be put on.

It is clear that History has a double function: it tells a story and it explains. It describes the events of the past in order to account for the conditions of the present. It uses yesterday to throw light on to-day. It analyses the complex fabric of modern civilisation by tracing it to the simpler beginnings from which it grew. It does all this and more. But primarily it tells a story, whether it is the story of the life of a man, or the story of the life of an idea; whether it traces the course of a movement or the development of a nation. It is always story-telling, and always story-telling with a purpose.

The tendency in the teaching of History—and in consequence the tendency in the testing of History—is in the same direction as the geographical trend. It is becoming less factual and more rational. In the earlier days the main stress was laid on names and dates and events; at the present time the main stress is laid on causal relationships. This may be illustrated by the treatment of wars. The older textbooks gave a dated and detailed list of battles but said little or nothing about their causes and their consequences. The newer textbooks have a great deal to say about causes and consequences and are silent about the mere incidents of warfare. To particularise a little further, the centre of interest tends to shift from military matters to political matters and from political matters to social matters. Kings and queens, unless they have had a hand in guiding the destinies of the nation, give place to the real makers of history.

Dates are a difficult problem. Some teachers would omit them altogether, preferring to have

History dateless, as others prefer to have Geography nameless. But a History without dates is like a road without milestones, or a realm without a landmark, or a sea with neither lightship nor shore. The objection is not to dates in themselves but to their multiplicity. If the milestones are too close together the road becomes as featureless as a paupers' graveyard. If the trees in a landscape are few and scattered they serve as landmarks; if they are many and close together they make the confusion of a forest. In finding our way in the vast realm of historic time we need a few Chanctonbury Rings: the New Forest is of little service. I have no doubt that every reading man has fixed in his mind a few dates to which he constantly refers the things of which he reads. The year 1809, for instance, is one of my own pet landmarks for the nineteenth century. It marks the birth of a galaxy of great men.[1] Gladstone, Tennyson, Darwin, Manning, and Lincoln are among the giants born in that year. Exactly a hundred years before, Dr. Johnson was born; and if he is to you the outstanding figure in the eighteenth century—if he is more interesting to you than the men who "hated boets and bainters"—then you will always find it illuminating to say: This happened when Johnson was so many years old.

We conclude that dates are good when taken in moderation. They steady the movements of our thoughts and give us a sense of historic proportion —a sense we sadly need; for we often fail to make

[1] The year 1812 was almost equally productive of "eminent Victorians."

allowance for temporal perspective, and we thus give point to the Autocrat's gibe: to-day's dinner subtends a larger visual angle than yesterday's revolution. We cannot do without dates. But they should be few, they should be prominent, they should be firmly fixed, and they should be constantly used. Advantage should be taken of some of the many coincidences which the curious may discover in History. To name but one: in 1588 there was an unsuccessful Catholic invasion of England; in 1688 a successful Protestant invasion.

Difficult as it is to determine the minimal contents of a course of study in geography, it is still more difficult in History. The field is larger; and it is steadily extending in two directions. New history is being made every day; old history is, in another sense, being re-made and is forcing back the borders of the prehistoric. In our own day a great war has dwarfed into insignificance every other war; and the discovery of an ancient tomb has illumined a little of the Egyptian darkness. It is the bewildering wealth of material that has forced us to become more parochial in our History than in our Geography. The schoolboy knows something about the geography of China: he knows nothing about the history of China.

It is possible that a scientific scheme of historical values may be arrived at by noting the frequency of reference to persons and events in current literature—in books and magazines and newspapers. Professor Bagley tried this plan, but gave it up. He came to the conclusion that such a study was suggestive, but was of doubtful value for establishing a final criterion. This is mere common sense.

Nobody who reads the serious magazines can fail to notice that a certain historical person will suddenly be much written about for no other reason than that somebody has recently published a book about him. Historical reference is subject to the fluctuations of fashion, and can only be taken as indicating an abiding public interest if the observations are extended over a long period of time. And, in any case, it would show obvious traces of national prejudice; and the scheme would thus entirely beg the question of outlook. It would tell us what the national outlook is, not what it ought to be. And that it ought to be different from what it is—wider and more humane, less the history of Englishmen and more the history of men—that it should be different from what it is, is what most of us have come to believe. From how narrow and parochial a point of view History is written can be clearly seen when we compare History as it is written in France, in Germany, or in America, with History as it is written in England. Even when both deal with the same events; nay, especially when both deal with the same events. A few years ago the Bishop of Manchester, in an address on the teaching of History, gave an account of impressions he formed as a boy of the relationship between England and France in the fifteenth and sixteenth centuries. He was taught that Henry V conquered France and attached it to the English throne. But he heard nothing more about our kingdom across the Channel until he was told the story of Mary weeping over the loss of Calais—the last of our French possessions. He wondered what had happened in the meanwhile. He could only suppose

that we had given the territory back to France in successive fits of generosity.

To the difficulty of choice and emphasis is added the difficulty of interpretation—of drawing sound inferences from the events that are studied. The fact that historians differ widely in the conclusions they draw from the same data—in the verdicts they give on the same evidence—is sufficient to show that there is no definite clear-cut interpretation of history. The network of cause and effect is so complex, and so little of it is open to view, that many of our inferences cannot but be personal and precarious. Still, there are certain obvious relationships about which there can be no difference of opinion, and certain lines of reasoning which are well within the capacity of school children. Among the attempts that have been made to test these simple relationships the best I know is the series of Pressey Tests in Historical Judgment. It consists of four tests, the first of which is a silent reading test, with a paragraph from a history of the United States as the material. The second test deals with time sequence and does not differ in purport from items 51–60 in my own History Test. The directions for the test are: "In each list on the page below there are the names of four men or of four events. In each list you are to draw a line under the name of the man who lived the longest time ago or of the event that occurred the longest time ago." I quote two of the 26 lists:

11. Stonewall Jackson. Pershing. Sheridan. Braddock.

20. Impeachment of Johnson. Panic of 1873. Civil War. Loss of National Bank's Charter.

The third test bears on the comparative importance of events. I quote the directions and two items: "In each of the lists below are the names of four events or of four men. In each list you are to draw a line under the most important event or man in the list."

6. Mad Anthony Wayne. Benedict Arnold. Patrick Henry. Isaac Putman.

26. Neutrality Proclamation. Postal Savings Banks. Federal Reserve Act. Demonitisation of silver.

The fourth test, which deals with cause and effect, appears in two "forms." I quote from Form A. Directions: "In each of the lists below there are three causes and one effect which was produced by the operation of the three causes. You are to find the effect and underline it."

10. Ku-Klux-Klan. Carpet-baggers. Dissension in Congress. Freedman's Bureau.

15. Anti-saloon League. Prohibition Amendment. Progress and Education. War-time needs.

I make similar quotations from Form B. Directions: "In each of the lists below there is one event in capitals and four events in small letters. You are to draw a line under the event in small type that was caused, either directly or indirectly, by the event in capitals at the beginning of the list."

3. CABOT'S DISCOVERIES: Voyage of Columbus. Founding of Quebec. Balboa's explorations. English claims to North America.

11. SQUATTER SOVEREIGNTY: Admission of Texas. St. Clair's Defeat. Fighting in Kansas. Assassination of McKinley.

I cite these tests, not for their subject-matter, which is manifestly unsuited for English children, but because they exemplify interesting ways in which the new technique may be applied.

The History examination at the end of this chapter has the same modest claims as the Geography examination at the end of the preceding chapter. It does not purport to be a standard test; still less does it pretend to say what a school-leaver should know about History. I have merely selected a certain limited period in the history of our own country—a period which may be safely assumed to have been studied in all English schools—and applied to that period the technique of the new examination. This test is suggestive and illustrative rather than prescriptive. At the same time, I have tried to invest it with as many of the merits of a standard test as I can compass. Although I have no doubt omitted much that is important, I have, I trust, included nothing that is trivial.

The things to do and the things to avoid in setting this examination are similar to those described in the preceding chapter for the Geography examination. The norms, too, are practically the same.

Forty-five minutes

HISTORY OF ENGLAND SINCE THE ACCESSION OF HENRY VII (TUDOR)

Fifteenth Century .	. 1485, 1492.
Sixteenth Century .	. 1509, 1534, 1588.
Seventeenth Century .	. 1649, 1660, 1666, 1688.
Eighteenth Century .	. 1707, 1757, 1783, 1789.
Nineteenth Century .	. 1801, 1805, 1815, 1832, 1846.
Twentieth Century .	. 1901, 1914.

When did the following events happen ? (The dates will be found among those given above.)

(1) The Battle of Waterloo.
(2) The Restoration.
(3) The Battle of Plassey.
(4) The Discovery of America.
(5) The French Revolution.
(6) The beginning of the Great War.
(7) The defeat of the Spanish Armada.
(8) The Battle of Trafalgar.
(9) The Execution of Charles I.
(10) The First Reform Bill passed.
(11) The Revolution and the Accession of William III and Mary.
(12) The American colonies became independent.

In each of the following sentences choose from among the names in brackets the one that will make the statement true.

(13) America was discovered by (Drake, Columbus, Cook, Raleigh).
(14) Shakespeare lived in the reign of (Henry VIII, Elizabeth, Anne, George III).
(15) The Reformation was begun on the Continent by (John Knox, Titus Oates, John Calvin, Martin Luther).

(16) Tobacco was introduced into England by (John Cabot, Thomas Cromwell, Sir Walter Raleigh, Roger Bacon).

(17) The Protector of England during the Commonwealth was (Cardinal Wolsey, Oliver Cromwell, Thomas Cromwell, Lord Burleigh).

(18) The Prime Minister of England just after the secession of America and during the time of the French Revolution was (William Pitt, Robert Walpole, Robert Peel, William Ewart Gladstone).

(19) The Battle of Trafalgar was won by (Wellington, Drake, Gordon, Nelson).

(20) At the Crimean War the nursing was organised by (Jane Austen, Florence Nightingale, Charlotte Brontë, Elizabeth Fry).

Supply the missing word or words where the number of the question appears in the following paragraphs.

England first became Protestant in the reign of (21), and finally in the reign of (22); but the separation from Rome took place in the reign of (23).

The people who wished to purify the usual practices of worship in the days of Elizabeth and James I were called (24). In the year 1620 some of them, known as the (25) sailed across the Atlantic to (26) in a ship called the (27).

The Stuart Period is noted for the struggle between (28) and (29). In the reign of Charles I civil war broke out, the King's soldiers being known as (30), and their enemies as (31).

The British first went to India as traders and formed in 1600 the (32) company. Their trade rivals there were the (33). The foundations of British rule in India were laid by Lord (34), who avenged the Black Hole of (35) by his victory at (36).

In the early years of the nineteenth century, England was engaged in a struggle with the French Emperor (37) who had practically become master of Europe. He was finally defeated at (38) by (39) and banished to (40).

Between the middle of the eighteenth century and the middle of the nineteenth century a great change took place in the main occupations of the people, a change known as the (41) Revolution. Instead of being engaged mainly in (42) the people became engaged mainly in (43), and they flocked from the (44) to the (45). The invention of (46) greatly increased the production of (47) goods, and added to our material (48).

After the Great War the League of (49) was formed for the purpose of preventing (50).

(51) Which of the following poets lived the earliest: Milton, Shelley, Shakespeare, Tennyson, Pope?

(52) Which of the poets in question (51) lived the last?

(53) Which of the poets named in question (51) occupied the middle position in regard to time?

(54) Which of the following statesmen lived the earliest: Peel, Burke, Gladstone, Wolsey, Strafford.

(55) Which of the statesmen named in question (54) lived the last?

(56) Which of the statesmen named in question (54) occupied the middle position in regard to time?

(57) Which was the earliest of the following events?

(*a*) The Trial of the Seven Bishops.
(*b*) The Franco-German War.
(*c*) The Execution of Mary, Queen of Scots.
(*d*) The French Revolution.
(*e*) The Great Fire of London.
(Name it by letter only, i.e. *a*, *b*, *c*, *d*, or *e*.)

(58) Which was the latest of the events mentioned in question (56)? Give the letter only.

(59) Which was the middle event of those mentioned in question (56) ?

(60) Who was the greatest admiral: Howard, Nelson, or Jellicoe ?

Some of the following statements are true and some are false. Read each carefully. If you think it is true put "Yes" on your answer paper; if you think it is false put "No." If you don't know put a dash ——.

I will now give three statements and show you how to deal with them.

(1) Henry VII was the first of the Tudor Kings. (1) Yes.

(2) Queen Elizabeth is the only queen who ruled over England. (2) No.

(3) George I knew more mathematics than George II. (3) ——.

Now answer the following in the same way. No statement must be considered true unless every bit of it is true.

(61) On the whole the Tudors were strong rulers and the Stuarts weak.

(62) The Reformation increased the influence of the Church in state affairs.

(63) The suppression of the Monasteries by Henry VIII proved of benefit to the poor.

(64) In the sixteenth century Spain was the greatest continental Power.

(65) In the eighteenth century France was the greatest continental Power.

(66) James I denied the Divine Right of Kings.

(67) Charles I ruled for many years without a Parliament.

(68) During the Commonwealth the country was ruled by the army rather than by parliament.

(69) Oliver Cromwell, by curbing the power of the Dutch at sea, increased England's sea-trade.

(70) The Restoration became possible because the army was popular and the Puritans beloved.

(71) Charles II preferred business to pleasure.

(72) James II was an ardent Protestant and tried to suppress the Roman Catholic religion.

(73) Since the "Glorious Revolution" the country has been governed by parliament through its ministers.

(74) The Bill of Rights made it illegal for Parliament to raise taxes.

(75) The Habeas Corpus Act was passed to prevent people being kept in prison without a trial.

(76) By the Act of Settlement the English sovereign was obliged to be a Roman Catholic.

(77) In the eighteenth century England was engaged in a long series of wars, mainly directed against the power of France.

(78) England and Scotland became one Kingdom in 1707.

(79) Sir John Moore secured British supremacy in Canada by the conquest of Quebec.

(80) The American colonies revolted because they objected to being taxed without their consent.

(81) Benjamin Franklin, the leader of the Revolutionaries in America, became the first President of the United States.

(82) Napoleon intended to invade England, but found it impossible because England had command of the seas.

(83) Napoleon's downfall began with his invasion of Russia and his disastrous retreat from Moscow in 1812.

(84) The wars with Napoleon caused low prices and much prosperity.

(85) The union of Ireland with England in 1800 caused great rejoicing in Ireland.

(86) In 1921 Ireland was granted Home Rule.

(87) The nineteenth century was a period of great political and social reforms.

(88) In the first half of the nineteenth century slavery was abolished, punishment for crime was made more humane, and religious liberty was extended.

(89) The means of travelling and of lighting houses and streets were unchanged during the whole of the Victorian era.

(90) The Repeal of the Corn Laws in 1846 made bread dearer and caused great distress.

(91) In 1870 education was made possible for all.

(92) Edward VII was called the War-maker.

(93) The Great War has set free several small nations.

(94) The Armistice which ended the fighting in the Great War was signed on the ninth hour of the ninth day of the ninth month of 1918.

(95) During the Great War there was a Coalition Government which included members of both the great political parties.

(96) Mr. Lloyd George was Prime Minister during the earlier part of the Great War and Mr. Asquith during the later part.

(97) Women have not yet received the parliamentary vote.

(98) England is now ruled by party government.

(99) As distinct from Liberals and Conservatives, a new political party, called the Labour Party, is rapidly increasing in number and influence.

(100) George V is the nephew of Queen Victoria.

KEY TO HISTORY TEST

1. 1815.
2. 1660.
3. 1757.
4. 1492.
5. 1789.
6. 1914.
7. 1588.
8. 1805.
9. 1649.
10. 1832.
11. 1688.
12. 1783.
13. Columbus.
14. Elizabeth.
15. Martin Luther.
16. Sir Walter Raleigh.
17. Oliver Cromwell.
18. William Pitt.
19. Nelson.
20. Florence Nightingale.
21. Edward VI.
22. Elizabeth.
23. Henry VIII.
24. Puritans.
25. Pilgrim Fathers.
26. America.
27. *Mayflower.*
28. Kings, the throne.
29. Parliaments.
30. Cavaliers, royalists.
31. Roundheads, republicans,
32. East India.
33. French.
34. Clive.
35. Calcutta.
36. Plassey.
37. Napoleon.
38. Waterloo.
39. Wellington.
40. St. Helena.
41. Industrial.
42. agriculture, farming.
43. industries, manufactures.
44. country.
45. towns.
46. machinery.
47. manufactured.
48. wealth, prosperity.
49. Nations.
50. War.
51. Shakespeare.
52. Tennyson.
53. Pope.
54. Wolsey.
55. Gladstone.
56. Burke.
57. *c.*
58. *b.*
59. *a.*
60. Nelson.
61. Yes.
62. No.
63. No.
64. Yes.
65. Yes.
66. No.
67. Yes.
68. Yes.
69. Yes.
70. No.
71. No.
72. No.
73. Yes.
74. No.
75. Yes.
76. No.
77. Yes.
78. Yes.
79. No.
80. Yes.
81. No.
82. Yes.
83. Yes.
84. No.
85. No.
86. Yes.
87. Yes.
88. Yes.
89. No.
90. No.
91. Yes.
92. No.
93. Yes.
94. No.
95. Yes.
96. No.
97. No.
98. Yes.
99. Yes.
100. No.

CHAPTER XX

GROUP TEST FOR JUNIORS

As the ordinary group test of intelligence presupposes the ability to read with ease, it is rarely given to children under ten years of age. When it is given it generally takes the form of a picture test in which the child is asked to add something to, or delete something from, a number of drawings, or in some other way to mark them or to modify them. I have much felt the need of a group test which includes the same variety of tasks as the ordinary written test, and yet can be applied to children of seven, eight, or nine years of age. Sometimes, indeed, there are whole schools that stand in need of such a test—schools in the slums, schools to which the homes are rather a hindrance than a help.

There is no ideal group test of intelligence any more than there is an ideal individual test of intelligence. The flaws of the Binet tests are to be found, often in an accentuated form, in every other scale that has been devised. And all the evidence goes to show that Binet's scale is strictly valid for those children only whose minds have, roughly speaking, been exposed to the same kind of influence in the home, and the same kind of influence in the school. It is not quite unaffected by environment; it is not quite independent of schooling. We can claim no more for the group tests than we can for the Binet tests: we can only claim that they measure a wider range of abilities and a more abiding type of aptitude than standard scholastic tests, which in their turn measure a wider range of

abilities and a more abiding type of aptitude than the ordinary examination.

I have already pointed out that we cannot, merely by means of mental tests, say how much of a slum child's mental disability is due to the stock from which he comes and how much is due to the home in which he dwells. The safest plan is to act at first on the assumption that the bulk of it is due to the home in which he dwells, and to watch the effect of remedying in the school the deficiencies of the home. But we cannot deal remedially with a child's disabilities until we know what they are. Indeed, there is only one thing that is initially more important than finding a child's disabilities, and that is finding his abilities. A good group test enables us to find both at the same time.

The less an intelligence test is affected by school studies the more fully does it serve its own specific purpose—the more completely and distinctively does it tell the teacher what nothing else can tell him. Judged by this standard the Chelsea Tests [1] come out well. Excellence of teaching does not seem to affect them to any appreciable extent. Where there are boys' and girls' schools in the same building, and the book-learning of the boys differs widely from that of the girls, there is never a corresponding disparity in the scores they get in the Chelsea Tests. In fact, there is no disparity at all: the scores are substantially identical. Teaching, if it tells at all, tells but slightly on these tests. I cannot say the same, however, of home environment —or rather of that unanalysed factor which seems to be compounded of home-environment and

[1] See *Group Tests of Intelligence*, chap. xiv.

heredity. For when I apply these tests to schools in sharply contrasted social surroundings I get results which stand abysmally apart. Let me illustrate. In the best elementary school—the best socially I mean—to which I have ready access the score in the Chelsea Tests is extraordinarily high. For the ages from 10 to 14 the averages are 35, 43, 41, 49, and 66 respectively. It will be observed that the children of twelve do worse than the children of eleven. This is accounted for by the fact that at the age of 11 a large contingent of the best scholars are drafted to secondary and central schools. Now note the results in a school at the opposite end of the social scale. The children of 10 could not even attempt the tests. The average scores for the age-groups from 11 to 14 were 7, 14, 17, and 21 respectively. The fourteen-year-old's in this school did not do as well as the ten-year-old's at the other school. Taken as a whole the pupils are five or six years behind in natural intelligence—or at any rate in their present capacity to deal with intellectual things, whether that capacity be natural or not. It is difficult to believe that such dullness is all due to parentage. One feels inclined to speak of it as Johnson spoke (rather unjustly in this instance) of the older Sheridan : " Why, sir, Sherry is dull, naturally dull ; but it must have taken him a great deal of pains to become what we now see him. Such an excess of stupidity, sir, is not in nature." In this ill-starred school the dullness of nature seems to be augmented by something else ; and that something else seems to be the home. It certainly is not the teaching ; for in scholastic attainments the disparity between the two schools

is not so marked. As judged by the results of the arithmetic and English tests given in this book the difference does not amount to more than three or four years.

These are two boys' schools; and it might be thought that the extraordinary contrast arises from certain extraordinary circumstances—certain unique factors within the walls of these schools. But it is not so. The findings have been confirmed at all points by applying the same tests to the corresponding girls' schools.

It is pure waste of time to apply the Chelsea Tests, or tests of equal difficulty, to children on so low an intellectual level as those in the poorer of the contrasted schools. I have accordingly constructed what I have called a Group Test for Juniors. It is suitable for the lower standards of all schools, and for all the standards of some schools. By "some schools" I mean those for which the ordinary group test is so difficult that it fails to secure a reasonable grading of the children.

In administering the Group Test for juniors the examiner does not provide the children with copies. He simply needs one copy for himself, from which he dictates the questions. He first gives each child a sheet of paper and directs him to prepare it for 100 very short answers. The key will suggest the amount of preparation desirable.

The examiner then says to the class: "I am going to give you some very easy questions to answer. Suppose I say: How many tails have five cats? What answer will you give? Five! That's right. Now try another. What is the colour of a tomato after it has been dipped in

water? Red! Quite right. The same colour as it was before it was dipped in water. Well, some of the questions you have to answer on your paper are just as easy as those. I am not trying to 'have' you, or to catch you: I am simply asking you common-sense questions. Do you understand?"

The examiner then proceeds to read out the questions one at a time, going through each twice before the children begin to write. The necessary instructions appear in detail on the test paper.

The words to be written on the blackboard should be placed one below the other in a vertical list. The purpose of writing them on the board is to remove the handicap from children who cannot spell. The number series, however, which are to be found in items 46, 47, 52, etc., should be written horizontally, as they appear in the test. The digits to be written down (see item 6) should be dictated at the rate of one per second without rhythm or emphasis.

To many of the questions there is no time-limit. The examiner notes when the children have finished (writing is a slow process with some of these young people) and then proceeds to the next question. When a time-limit is mentioned, it refers, as a rule, to the period for which the words or numbers are allowed to remain on the blackboard after the examiner has ceased speaking.

I am indebted for the following norms to the teachers in the County of Kent, who have applied the test to 14,037 children. 6 + means all the six-year-olds, 7 + all the seven-year-olds, etc.

Age	6+	7+	8+	9+	10+	11+	12+	13+
Norms	20·5	41·7	55·5	65·7	72·3	77·6	80·6	83·1

GROUP TEST FOR JUNIORS

Instructions to the examiner are printed in italics.

The children are not given copies of this test. The examiner dictates the questions.

Each question is to be read twice except when numbers or letters are dictated.

Words in brackets have to be written on the blackboard and read to the class as well.

The time placed after the brackets indicates the time the words should remain on the blackboard after the examiner has ceased speaking.

There is no time limit for the writing of the answer.

1. How many legs has a three-legged stool ?
2. From black sheep we get black wool. What is the colour of the milk we get from a black cow ?
3. (Bat, doll, gloves, scooter.) 15 *seconds.*
   Which of these things would you choose as a present for your mother ? A bat, a doll, a pair of gloves, or a scooter ? Write one word only.
4. If hard is the opposite of soft, what is the opposite of wet ?
5. (Daisy, snowdrop, rose, lion, daffodil.) 15 *seconds.*
   Daisy, snowdrop, rose, lion, daffodil. Four of these words mean the same sort of thing, but one means something different. .Write down the word that means something different.
6. *Read the digits once only, at the rate of one per second.*
   I am going to read some numbers. When I have finished write them down : 2, 7, 4, 5.
7. (Cat, dog, book, horse, sheep.) 15 *seconds.*
   Cat, dog, book, horse, sheep. Four of these words mean the same sort of thing, but one means something different. Write down the word that means something different.

8. (Richard, Tom, David.) 15 *seconds.*
At a party Richard ate more than Tom and David ate more than Richard. Who ate the least ?

9. What day comes before Saturday ? Write only the first letter of the word.

10. (Herring, mackerel, wolf, trout, salmon.) 15 *seconds.*
Four of these words mean the same sort of thing, but one means something different. Write down the word that is different in meaning.

11. (Mary, Jane, Anne.)
Mary is older than Jane, and Anne is younger than Jane. Who is the oldest ?

12. A boy looked through a gate into a field, the whole of which he could see, and he saw six sheep grazing. His sister also looked through the gate and she saw six sheep grazing. How many sheep were there in the field ?

13. What is the opposite of asleep ?

14. *Explain the meaning of epitaph.* (True, false.)
Is this epitaph likely to be true or false ? Here lies the body of John Pound, who was lost at sea and never found.

15. Two tramps are 12 miles from London. How many miles will each have to walk to get to London ?

16. What is the smallest number of match-sticks you will need for making a square without breaking any of the sticks ?

17. What is the opposite of cheap ?

18. What is the opposite of up ?

19. What is the opposite of narrow ?

20. (No pockets, no sense, handier.)
Why do people wear watches on their wrists ? Because they have no pockets, or because they have no sense, or because the watches are handier there ?

21. (4 yards, 4 miles, 40 miles, 400 miles.)
How far can a man walk in an hour ?

22. *Say the word London and spell it.* Write down the two middle letters of the word L-o-n-d-o-n.

23. (Pipe, bicycle, cigars, muff.) 15 *seconds.*
If your father did not smoke, which of the following would you choose as a present for him ? A pipe, a bicycle, a box of cigars, or a muff ?

24. *Explain on the blackboard by another example how to deal with a dissected sentence.* (Clock time a the tells.) 15 *seconds.*
Turn these into a proper and sensible sentence and write down the last word in that proper sentence.

25. (Drain, view, accidents, pretty.)
Why are the roads high in the middle and sloping towards the sides ? Is it in order that the rain may drain off or that drivers may have a good view, or to prevent accidents, or because it looks pretty ?

26. *Dictate the digits once only at the rate of one per second.* Write the figures, 4, 8, 1, 9, 2.

27. *Explain what is required by a preliminary example.* 20 *seconds.*
What is the next number in this row of figures 9, 8, 7, 6 . . . ?

28. (Head, feet, hands.)
Do the people on the other side of the world stand on their heads, feet, or hands ?

29. (Buy, tell, give.)
If the grocer gives you too much change what should you do ? Buy him some chocolate, tell him of his mistake, give the money to your mother ?

30. Write down the middle word in this sentence : Simple Simon met a pieman.

31. (Green, cheap, natural.)
Why is grass a good food for cows ? Because it is green, because it is cheap, or because it is their natural food ?

32. (Happy, pleasant, gay, sad, cheerful.) 15 *seconds.*
Four of these words mean the same sort of thing, but one means something different. Write down the word that is different in meaning.

33. How many legs has an ostrich ?

34. (Instrument, boy, stone, plant).
Write down the word which best describes what a herb is.

35. *Dictate the letters once only at the rate of one per second.*
Write down the letters f, h, p, t, r.

36. (Bread, meat, potatoes, water, cheese.) 15 *seconds.*
Four of these words mean the same sort of thing, but one means something different. Write down the word that is different in meaning.

37. (Mother, father, aunt, sister, niece.) 15 *seconds.*
Four of these words mean the same sort of thing, but one means something different. Write down the word that is different in meaning.

38. (Mice cats catch.) 15 *seconds.*
Turn these words into a proper sentence and write down the first word in that proper sentence.

39. (Little a Tom sister has.) 15 *seconds.*
Turn these words into a proper sentence and write down the last word of that sentence.

40. (Carts, taxis, horses, houses, smoke.) 15 *seconds.*
Write down the word which says what a town always has.

41. (Green, red, black, yellow, blue.) 15 *seconds.*
Four of these words mean the same sort of thing, but one means something different. Write down the word that is different in meaning.

42. Write the figures 6, 3, 5, 0, 7, 2.

43. What number comes next but one after 15 ?

44. (Wednesday, Friday.)
Which is the nearer to a Sunday ? Wednesday or Friday ?

45. (Bees, sugar, hives, flowers.)
Where does honey come from originally ? From bees, from sugar, from bee-hives, or from flowers ?

46. *Explain by means of a preliminary exercise.* (10, 8, 6, 5, 4, 2). 20 *seconds.*
Write down the number that is wrong and ought not to be there.

47. (3, 6, 7, 9, 12, 15.) 20 *seconds.*
Write down the number that is wrong and ought not to be there.

48. Write the letters o, e, w, i, r, m.

49. (Grass, plum, leaf, tree, walnut.) 15 *seconds.*
Think of the way in which these things are alike, apple, pear, cherry. Then find the one thing named on the board that is most like them.

50. (Sat a three tree crows on black.) 15 *seconds.*
Turn these words into a proper sentence and write down the first word in that proper sentence.

51. (Was a John boat out taken in.) 15 *seconds.*
As in 50.

52. (5, 10, 15, 20.) 20 *seconds.*
What is the next number in this row ? 5, 10, 15, 20 ?

53. If a candle takes two hours to burn away how many hours will it take two candles if they are lighted at the same time ?

54. (2, 4, 5, 6, 8, 10.) 20 *seconds.*
Write down the number that is wrong and ought not to be there.

55. (8, 9, 7, 6, 5, 4.) 20 *seconds.*
As in 54.

56. (21, 64, 15, 39, 42.) 30 *seconds.*
Arrange these numbers in your mind in order of size and write down the middle one.

57. (Possible, impossible.)
John Smith lived in four different towns one after the other, remaining ten years in each. Is that possible or impossible ?

58. (1, 2, 3.)
A boy tried three times to explode a Christmas cracker. Which time did it explode ? The first, second, or third ?

59. Write the figures 1, 4, 7, 3, 9, 6, 0.

60. (Ribbon, fur, mouse, milk). 15 *seconds.*
Write down the word which says what a cat always has.

61. (1, 3, 5, 7, 8, 9.) 20 *seconds.*
Write down the number that is wrong and ought not to be there.

62. *Allow* 30 *seconds.*
In a street of houses all the same size the odd numbers are on one side and the even numbers on the other, and both sets of numbers begin from the same end of the street. What number is opposite No. 6 ?

63. (Down of first alphabet the put letter the.) 15 *seconds.*
Turn this into a proper sentence and do what it tells you.

64. (2, 4, 6, 8 . . . .) 20 *seconds.*
What is the next number in this row of figures ? 2, 4, 6, 8 . . . ?

65. (Leather, wood, seat, carving.) 15 *seconds.*
Write down the word which says what a chair always has.

66. (Paper on cross a your put.) 15 *seconds.*
Turn this into a proper sentence and do what it tells you.

67. (Boys . . . do wrong should be punished). 15 *seconds.*
Write down the missing word.

68. (1, 2, 3, 4, 8, 5.) 20 *seconds.*
Write down the number that is wrong and ought not to be there.

69. (Likely, possible, impossible.)
It began to rain yesterday morning and kept on without stopping for three days. Is this likely, possible, or impossible?

70. (1, 3, 5, 7 . . .) 20 *seconds.*
What is the next number in this row of figures?
1, 3, 5, 7 . . . ?

71. (March, May, July, October.) 15 *seconds.*
Which is the hottest of the four months? March, May, July, and October?

72. (Possible, impossible.)
Mr. Robinson lived in three different towns one after the other, remaining in each of them three years longer than in either of the others.

73. (October, January, March, April.) 15 *seconds.*
Which is the coldest of the four months? October, January, March, and April?

74. (Frock, ball, cake, sweets, pictures.) 15 *seconds.*
Write down from the blackboard the one word that is most like these three in meaning: top, bat, marbles.

75. (Lead, feathers, same.)
Which is heavier, ½ lb. of lead or 1 lb. of feathers?

76. (Sky, clouds, faint, moved.)
Why don't we see stars in the day-time? Because they are behind the sky, or because the clouds have hidden them, or because they are fainter than the sun, or because they have moved to the other side of the earth?

77. (Fish, bird, reptile, insect.)
Is a lizard a fish, a bird, a reptile, or an insect ?

78. (Wood, rain, tree, tar, fire.) 15 *seconds.*
Write down from the blackboard the one word that is most like these three in meaning: coal, ink, soot.

79. How many halfpenny buns are there in a dozen ?

80. (Laughter, smiles, excitement, joy.) 15 *seconds.*
Write down the word which says what happiness always has.

81. (Misery, idleness, happiness, sleep.) 15 *seconds.*
Write down the word which says what laziness always has.

82. Write down the last letter but one in the last word but one in the sentence "Mary had a little lamb."

83. (Front, hind, same.)
In each of the front wheels of a four-wheeled cart there are 16 spokes, and in each of the hind wheels there are 12 spokes. When the cart moves which spokes move the faster ?

84. (String, tape, elastic, same, O.)
Which is the largest ? A piece of string, a piece of tape, or a piece of elastic ? If you think they are the same put "Same"; if you can't say without seeing them put "O."

85. (Book, head, house, stick, tie.) 15 *seconds.*
Write down from the blackboard the one word that is most like these three in meaning: hat, coat, boots.

86. (8, 8, 6, 6 . . .). 20 *seconds.*
What is the next number in this row ? 8, 8, 6, 6 . . . ?

87. (Goddess, queen, fairy, singer.)
Write down the word which best describes who Psyche was.

88. What relation to me is my mother's sister's son ? Spell the word as best you can. No marks will be taken off for bad spelling.

89. (Chalk, coal, same, O.)
Which is the larger ? A lump of chalk, or a lump of coal ? If you don't know without seeing them put down "O."

90. (Plum, apple, apricot, peach, cherry.) 20 *seconds.*
Four of these words mean the same sort of thing, but one means something different. Write down the word that is different in meaning.

91. (Good, bad.)
Is this good advice or bad ? Potatoes should always be boiled in cold water.

92. (1, 2, 4, 8, 10, 16.) 20 *seconds.*
Write down the number that is wrong and ought not to be there.

93. (Song, poem, tale, novel.)
Write down the word which best describes what a sonnet is.

94. A family of children consists of three brothers, each of whom has only one sister. How many brothers and sisters are there altogether ?

95. What relation to me is my sister's daughter ?

96. (1, 2, 4, 8 . . . .) 20 *seconds.*
What is the next number in this row ? 1, 2, 4, 8 . . . ?

97. (1, 2, 4, 7 . . . .) 30 *seconds.*
What is the next number in this row ? 1, 2, 4, 7 . . . ?

98. How many great-grandmothers have you alive or dead ?

99. (Prison, fish, anger, museum.)
Write down the word which best describes what dudgeon means.

100. (Metallic, happy, woody, sunny.)
Write down the word which best describes what silvan means.

KEY TO GROUP TEST FOR JUNIORS

1. 3.
2. White.
3. Gloves.
4. Dry.
5. Lion.
6. 2745.
7. Book.
8. Tom.
9. F.
10. Wolf.
11. Mary.
12. 6.
13. Awake.
14. False.
15. 12.
16. 4.
17. Dear.
18. Down.
19. Wide.
20. Handier.
21. 4 miles.
22. nd.
23. Bicycle.
24. Time.
25. Drain.
26. 48192.
27. 5.
28. Feet.
29. Tell.
30. Met.
31. Natural.
32. Sad.
33. 2.
34. Plant.
35. fhptr.
36. Water.
37. Father.
38. Cats.
39. Sister.
40. Houses.
41. Black.
42. 635072.
43. 17.
44. Friday.
45. Flowers.
46. 5.
47. 7.
48. oewirm.
49. Plum.
50. Three.
51. John.
52. 25.
53. 2.
54. 5.
55. 9.
56. 39.
57. Possible.
58. 3.
59. 1473960.
60. fur.
61. 8.
62. 5.
63. A.
64. 10.
65. Seat.
66. X.
67. Who.
68. 8.
69. Impossible.
70. 9.
71. July.
72. Impossible.
73. January.
74. Ball.
75. Feathers.
76. Faint.
77. Reptile.
78. Tar.
79. 12.
80. Joy.
81. Idleness.
82. 1.
83. Same.
84. O.
85. Tie.
86. 4.
87. Goddess.
88. Cousin.
89. O.
90. Apple.
91. Bad.
92. 10.
93. Poem.
94. 4.
95. niece.
96. 16.
97. 11.
98. 4.
99. Anger.
100. Woody.

CHAPTER XXI

SCRIPT WRITING

HANDWRITING is one of the things with which the new examiner cannot cope. Taken as a whole it is too complex a product. And even when analysed—when considered in its separate aspects of speed, legibility, and beauty (the three aspects which are generally viewed as examinable) it fails to present a front over which the new examiner can pass his measuring-tape. Except in the matter of speed. The number of letters written per minute is easy to count and free from ambiguity. But speed in itself is of little import. It is no great comfort to the reader of an illegible manuscript to learn that it was written in record time. Next to speed comes legibility. It seems a sound procedure to compare the legibility of two scripts of equal length by noting the length of time it takes to read them. If one takes twice as long as the other, the other is twice as legible as the one. But this plan suggests an inevitableness in the scoring which does not in point of fact exist. Another reader might conceivably reverse the decision of the first. And even when the same reader marks all the scripts his interest may flag and his attention fluctuate in sufficient measure to

mar his results and vitiate his findings. The method, in fact, is not fool-proof.

It is when, however, we try to measure the third aspect, beauty, that we come to realise the full hopelessness of our task. That there is an objective standard of beauty in penmanship as in everything else my own philosophy forces me to believe; but at the same time that homelier philosophy which we call common sense convinces me that this objective scale of values is, at present, at any rate, beyond our reach. So dimly is it seen, so vaguely is it apprehended, that individual judgments are far too erratic to satisfy the demands of a scientific scheme of marking.

The best device that has yet been found for steadying judgments based on general impression is the Thorndike method of samples. Samples of handwriting already marked are arranged to form a scale, and any specimen that has to be assessed is compared with the scale and given the score of the sample which it most closely resembles. It is not a perfect method, but it is the best we have. And it is certainly more satisfactory as applied to penmanship than as applied to the other two subjects for which similar scales have been made.

Every nation has its own characteristic handwriting. America has hers; and she should be allowed to keep it. It resembles, with a few subtle points of difference, the business or civil service hand which was popular in our schools ten years ago and is still the approved style in the world of commerce. Its distinguishing feature is its emphasis on loops and links, its insistence on padding—on the very parts that reduce the legibility of the

manuscript and destroy its beauty. They reduce its legibility because they constitute the factors common to all the letters, not the factors peculiar to each. They destroy its beauty because they screen the essential fabric with accidental details. At any rate we in England have definitely made a new departure in the teaching of penmanship. In most of our schools we have discarded pot-hooks and hangers as preliminaries, in favour of a simple form of the printed word. The advantages of the script system, as it has come to be called, are enormous. The youngest of beginners take kindly to it and soon get to write it with ease. Almost from the very outset handwriting becomes a genuine instrument of expression. For the first time in the history of the infant school the children write quite long pieces of free composition, and write them with the same charming spontaneity which they display in their drawings. And as it has proved a liberating factor in the kindergarten, freeing the pupils from the fetters of a formal cursive hand, so has it proved a humanising factor in the schools for older pupils, imparting to them something of that peculiar culture that comes through fine art and fine craftsmanship.

There is often, however, a tendency in the senior school either to continue unchanged the primitive script of the infant school or to change it for the old-fashioned cursive hand. Neither course is to be commended. A far better plan is to evolve a cursive hand out of formal script. To repeat what I have said elsewhere[1]: "If script is to fulfil its

[1] The Introduction to *Modern Script for Schools*, by T. E. Raw, p. 14.

proper function in the school, if it is to be a real reform which will distinctly raise the general level of the nation's handwriting, it must be taught as a progressive system. It must follow a definite course. It must not rest as the baby writing of the infant school; nor even as the decorative calligraphy of the art school. Life is too short to permit of our writing to our friends or our clients as though we were presenting them with illuminated addresses. And the examinee who affected the formal manuscript style would pass in handwriting and fail in everything else. The writing ultimately reached must look like writing, and not like a bad imitation of print. It must have character and depart from the stark neutrality of the straight line and the circle. In fact, though it begins as printing it must end as writing."

It was with these facts in mind that I drew up the appended Scale for Script Writing. The order is based on the combined judgments of several examiners—all of them people whose opinions are entitled to respect. They include teachers who hold the Art Master's Certificate, and heads of schools who have taken a leading part in developing the script system of writing.

To judge a specimen of handwriting all the examiner has to do is to find among the samples in the scale the one to which it most closely approximates in merit, and then to give it the same score as that sample.

10 MARKS

I am fifteen years of age, and since leaving school I have attended The Premier Business College for one year, during which I have

Thousands of pilgrims and priests visit these spots every year at stated festivals and thousands more unable

10 MARKS

" Perfect ensemble singing is all the more certain this year, because musical intelligence and technique generally are far superior to what they were even three years ago. The opportunities for musical

Thousands of pilgrims and priests visit these spots every year at stated festivals; and thousands more unable to do so, receive from their

9 MARKS

Thousands of pilgrims and priests visit these spots every year at stated festivals; and thousands more, unable to do so, receive from their priests

I am fifteen years of age, and since leaving school I have attended The Premier Business College for one year, during which I have studed commercial

Then softly Wendy called to her mother. But when Mrs. Darling looked round she simply couldn't believe that the children were

"Perfect ensemble singing is all the more certain this year, because musical intelligence and technique generally are far superior to what they were three years ago. The opportunities

8 MARKS

The tree tops are soft as velvet, and in the evening at twilight are all bejewelled with tiny mauve, and white, and blue

2 Young Milton was fortunate in his up-bringing, His Father loved poetry and music, and was able to give his son the very best education that the age could provide.

8 MARKS

In reply to your advertisement in today's Journal for a Junior Clerk, I beg to offer you my services.

I am fifteen years of age, and

We feel justified in concluding that you would not have communicated with us had you not had in your mind the purchase of a typewriter

7 MARKS

Here is an excellent example of an expansion written by Froude the historian. The idea is, "In this way sufficient food was obtained to support

Here is an example of an expansion written by Froude the historian. The idea is "In this way sufficient food was obtained to support the

7 MARKS

"Perfect ensemble singing is all the more certain this year, because musical intelligence and technique generolly are far superior to what they were even three years ago. The opportunities for musical education

The quality of mercy is not strain'd,
It droppeth as a gentle rain from heaven
Upon the place beneath: it is twice blest;
It blesseth him that gives and him that takes:

6 MARKS

In those days there was only one kind of church in England, and everybody went to it. Each village had one church, and even a town

We feel justified in concluding that you would not have communicated with us had you not had in your mind the purchase of a typewriter.

6 MARKS

The tree tops are soft as velvet, and in the evening at twilight are all be jewelled with tiny mauve, and white and blue

In her lap the princess was to hold three golden apples, and the man who could take them from her was to have

5 MARKS

It blesseth him that gives and him
that takes.
'Tis mightiest in the mightiest: it
becomes the throned monarch better

Here is an excellent example,
of an expansion written by
Froude the historian. The idea
is:- "In this way sufficient

5 MARKS

Eight years before England lost Shakespeare she gained another great poet, who gave to our language some of the most exquisite and

Ye Mariners of England
That guard our native seas!!
Whose flag has braved a thousand years
The battle and the breeze!

4 MARKS

The next day, just as the two brothers were again about to set off, Boot begged hard that he might go with them. They, however,

"Perfect ensemble singing is all the more certain this year because musical intelligence and technique generally are far superior to what they were even three years ago.

The monks landed in Kent
Now the King of Kent
was a wise man. He
liked the monks, and he

Once upon a Time a poor
little girl lived with her
Stepmother. Her father and mother
were dead, and she was very badly used

2 MARKS

Do you hear the kettle ?" said
the Saucepan to her friend
the Frying Pan. "She never
Seems to be happy unless She

boxed his little saucy ears,
And then away I strode;
But, since, I've found that weary path
Is quite a common road.

1 MARK

Ye Mariners of England
That guard Our native seas!
Whose flag has braved a thousand years
The battle and the breeze!

"Do you hear the kettle said the Sauce pon
to her friend the frying pon. She never seems
to be happy unless she is making a
noise!" "I am tired of hearing her

INDEX